Urban-scale Modelling for WesCon

January 2025

Kirsty Hanley, Humphrey Lean

Technical Report 667

https://doi.org/10.62998/DITE4759

OFFICIAL-SENSITIVE

© Crown Copyright 2025, Met Office

Document history

Version	Purpose	Date
1.0		27 January 2025

Prepared by

Kirsty Hanley, Senior Scientist, January 2025

Reviewed by

Humphrey Lean, Science Manager, January 2025

Authorised for issue by

Huw Lewis, Strategic Head, January 2025

© Crown copyright 2025, Met Office

Contents

Document history....2
Contents....3
1. Executive Summary....4
2. Introduction....5
3. Model Description....7
3.1. The WMV ensemble....7
3.2. Higher resolution ensembles....9
3.3. Science configurations....11
3.4. Verification metrics....13
4. Routine running results....15
4.1. Comparison of the WMV with MOGREPS-UK....15
4.1.1. Case 1 (IOP 8): Post-frontal showers aided by surface convergence 16
4.1.2. Case 2 (IOP 11): Organised deep convection....18
4.1.3. Case 3 (non-IOP): Diurnal shallow convection....22
4.2. FSS Verification....23
5. Post-campaign ensemble simulations....26
5.1. IOP 5: Diurnal convection associated with low-level convergence....26
5.2. IOP 6: Diurnal shallow convection....32
6. Conclusions....41
References....43

© Crown copyright 2025, Met Office

1. Executive Summary

Extended evaluation of the variable resolution 300 m Wessex Model (the “WMV”) was conducted through a daily running trial during summer 2023, running as an ensemble nested inside the operational MOGREPS-UK 2.2km ensemble. Overall, the WMV looks promising for high-impact convective events as it is better able to represent the organisation of convection into lines or larger storms whereas MOGREPS-UK tends to simulate isolated, circular storms. This often leads to more reliable probabilities of heavy rainfall in the WMV ensemble compared to MOGREPS-UK. Fraction Skill Score analysis showed that the WMV is more skilful at smaller neighbourhoods than MOGREPS-UK with the displacement between observed and modelled precipitation being less in the WMV than in MOGREPS-UK.

However, there is still an issue with the WMV producing too many small precipitating showers in situations where there should only be shallow clouds. This is thought to be a result of shallow clouds getting too deep in the model, possibly due to the updrafts not entraining enough dry air. Going to finer horizontal resolution alleviates this issue slightly. At 100m grid length, the updrafts (and hence the clouds) do not penetrate as deep even though they are stronger. This suggests the entrainment processes are better resolved at this grid length. Detailed comparisons with the WesCon turbulence and vertical velocity observations will hopefully allow us to better understand this issue. Additionally, it may be possible to improve the performance of the 300m ensemble by using new parameterisations, such as the prognostic turbulent kinetic energy scheme or the scale-aware convection scheme, currently being developed at the Met Office.

The WesCon field campaign, which took place during summer 2023, has provided a wealth of observational data that will be used to support new process understanding and improving our km and sub-km predictions in UM and LFRic contexts, as well as underpinning broader aspects of model physics development. This work will be carried out in conjunction with the ParaChute programme, a joint Met Office and NERC project to improve the representation of turbulence in km and sub-km scale models. Additionally, the WMO 2024 Paris Olympics Research Demonstration Project has provided unique observations and international model intercomparison of urban-scale models. During summer 2024 we routinely ran a 300m ensemble for the Paris Olympics to form part of the pathway towards implementation of a 300m ensemble “trailblazer” system on the next HPC.

© Crown copyright 2025, Met Office

© Crown copyright 2025, Met Office

2. Introduction

The Urban-scale Modelling Project, which is part of the Path to High Resolution R&I theme, aims to utilise our increased supercomputing capacity to enable the Met Office to develop the capability for running forecast models at O(100m) grid length to provide information at increasingly local scales. One of the objectives of the Urban-scale Modelling project is to deliver an upgraded 300m London model.

The London model is a deterministic model, one-way nested in the UKV. It is run out to 36 hours twice daily at 06Z and 18Z. While fog forecasts from the London model have been shown to have improved performance compared to the UKV (Boutle et al., 2016), previous work looking at convection in the London model has demonstrated that while certain aspects of the convection are improved (such as the distribution of rainfall rates and the initiation time of convection), the storm structures are very similar to those in the UKV suggesting the London model domain is too small to fully allow convection to spin-up (Hanley and Lean, 2024). To combat this, Hanley and Lean (2024) tested a new variable resolution 300m London model (LMV), which allows the lateral boundaries of the domain to be moved further away from the areas of interest at much cheaper computational cost than running a much larger fixed grid length 300m model. Based on a small number of cases, the LMV was shown to produce surface precipitation features comparable to a large fixed-grid-length 300m model making variable resolution a viable option for increasing the size of the London model domain without increasing the computational costs too much.

Convective-scale predictability is severely limited by the inherent predictability of the atmosphere and by errors inherent in NWP models. As a result, we would not expect a storm to be in the same place in the model as it is in observations. This problem is amplified with a small domain, such as the London model domain, as storms can appear missed altogether. As a result, for the London model upgrade an urban-scale “trailblazer” forecast system comprising of a variable resolution 300m London model ensemble is proposed for implementation on the Met Office’s Gen1 HPC with the timescale for operational implementation being 2025. Ahead of this, several trial periods for the urban-scale trailblazer system were proposed, running over London in summer 2022, running over Wessex in summer 2023 to coincide with the WesCon field campaign and running over Paris in summer 2024 to coincide with the Paris Olympics (Figure 1).

© Crown copyright 2025, Met Office

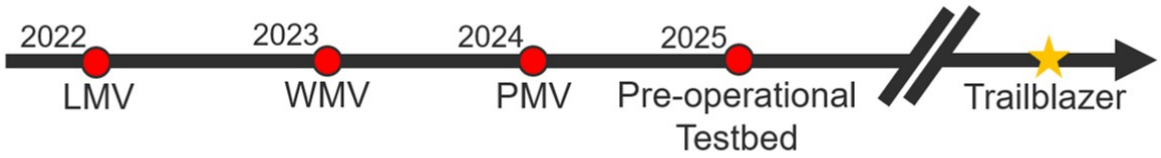


Figure 1. Timeline towards implementation of the urban-scale trailblazer forecast system.

Extended evaluation of the LMV was conducted through a daily running trial during summer 2022, running as an ensemble nested inside the operational MOGREPS-UK 2.2km ensemble (Hanley and Lean, 2024). Overall, the LMV performed well for several high-impact convective events as it was better able to represent the organisation of convection into lines or larger storms whereas MOGREPS-UK tends to miss the organisation and simulate isolated, circular storms. The LMV also performed well for several high-impact heat wave events, providing extra local detail for near-surface temperatures. However, there is still an issue with the LMV (and urban-scale versions of the UM in general) producing too many small precipitating showers in situations where there should only be shallow clouds.

The Wessex Summer Convection Experiment (WesCon) was a field programme that took place in summer 2023 and was designed to provide detailed observations of summertime convection over the UK. Measurements were made of both the pre-convective environment and the resulting convective, dynamical and cloud structures and precipitation fields. In-situ thermodynamic and microphysical airborne measurements made using the FAAM research aircraft were co-located with comprehensive radar observations (Chilbolton and Met Office Wardon Hill research Radar), to provide a detailed view of the three-dimensional structures of clouds and precipitation fields, and of convective organisation across a range of mesoscale conditions. Additional ground measurements, including a super-site, provided detailed boundary layer observations, remote sensing capability and radiosondes. The measurements were made at high enough resolution and over sufficient spatial scale to challenge O(100m) and O(1km) grid length models.

To benefit from the availability of high-quality, co-located observations, during summer 2023 a variable resolution 300m Wessex model (WMV) was run daily as an ensemble nested inside the operational MOGREPS-UK as described in Section 3. Results from the WMV trial, including Fraction Skill Score verification are presented in Section 4 of this report. Post-campaign, 200m and 100m grid length ensemble simulations of several WesCon Intensive Observation Periods (IOPs) were performed to understand the resolution dependence of convection. These results, including initial comparison with WesCon flight observations are presented in Section 5.

Additionally, the Summer 2023 UK Testbed (which ran alongside WesCon)

© Crown copyright 2025, Met Office

included a post-event activity to 1) evaluate the convection characteristics of the WMV (Flynn et al., 2023) and 2) to assess the added value an urban-scale ensemble provides for forecasting in urban areas (Glazer et al., 2023). Feedback from testbed participants (which included several Operational Meteorologists) was largely positive, specifically the WMV ensemble was viewed as giving a useful indication of storm structure and organisation, which is often lacking in MOGREPS-UK, with the greater probabilities of extreme rainfall providing a better steer of high impact weather. However, in shallow convection regimes testbed participants found the WMV too eager to produce widespread small precipitating showers. It was also noted that on a few occasions the WMV ensemble produced erroneously high 10m wind gusts associated with convective precipitation.

© Crown copyright 2025, Met Office

3. Model Description

During summer 2023, the Urban-scale Modelling Group within the Regional Modelling Evaluation and Development (RMED) Team ran a number of experiments both routinely and post-event. These are described in more detail in the following sections and include:

1) A routinely running 18-member variable resolution 300m “WMV” ensemble
2) Post-event 18-member variable resolution 200m and 100m ensembles

3.1. The WMV ensemble

An 18-member variable resolution 300m Wessex model (WMV) ensemble was run daily during the period 9 June 2023 to 9 September 2023 with initial conditions (ICs) and lateral boundary conditions (LBCs) provided by the operational MOGREPS-UK. The WMV domain (shown in Figure 2) consists of three regions: a high-resolution fixed-grid-length inner region, a variable resolution stretching zone and a course-resolution outer region. The inner fixed-resolution part of the domain has a grid length of 300m. The variable resolution stretching zone (indicated by the region between the red line and the black dashed line) uses an inflation factor of 1% to increase the grid length from 300m to 1.5 km. Beyond the stretching zone is a courser-resolution 1.5 km region. Due to the way the grid is stretched, the grid length in this region will be 1.5 x 1.5 km (corner regions), 1.5 km x 300 m (east/west of stretching zone) or 300m x 1.5 km (north/south of stretching zone). A comparison of the WMV and MOGREPS-UK is given in Table 1.

© Crown copyright 2025, Met Office

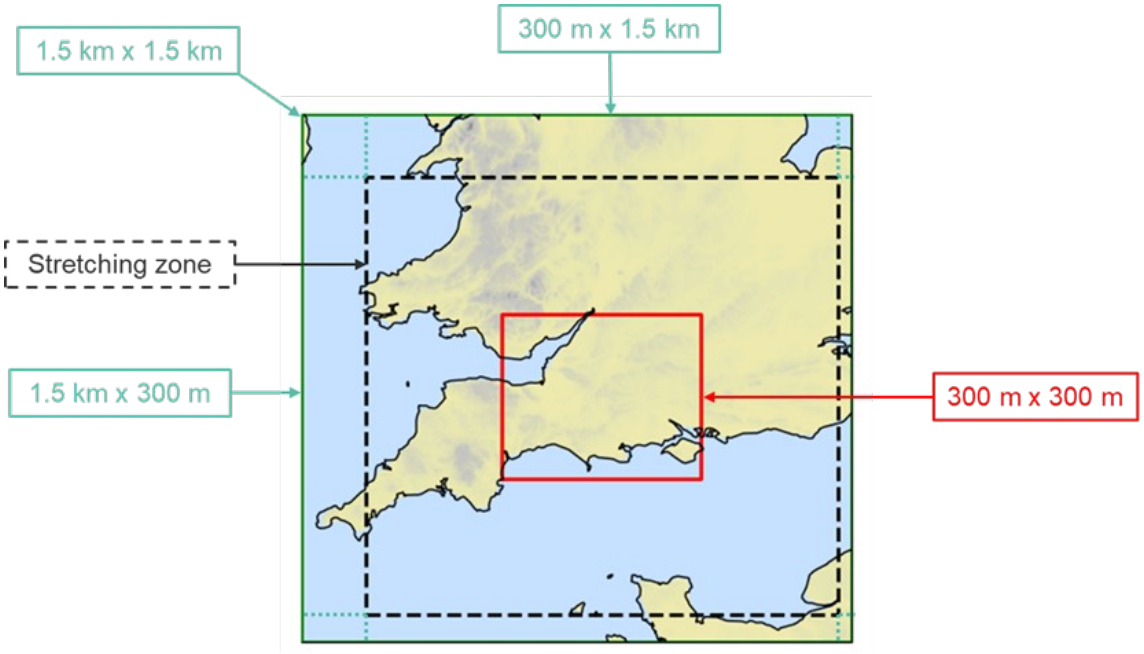


Figure 2. The variable resolution 300m Wessex model (WMV) domain. The red line shows the inner 300-m grid length region, the region between the red line and the dashed black line shows the stretching zone and beyond the dashed black line is the 1.5-km grid length region.

Model	Grid length (km)	Grid points	Levels	Timestep (s)	Science configuration
WMV	0.333 -> 1.5	856 x 774	70	12	RAL3.1
MOGREPS-UK	2.2 -> 4.4	740 x 752	70	100	RAL2-M

Table 1. A comparison of the WMV and MOGREPS-UK domains.

MOGREPS-UK is a variable resolution ensemble with 2.2 km grid space in the inner region, extending to 4 km at the edges. Since March 2019 it has been run as a time-lagged ensemble with 3 members initialised every hour and run out to T+120 (Porson et al., 2020). Every hourly cycle, all 3 members are centred on the high-resolution deterministic UKV analysis, derived from the 4D-Var Data Assimilation system. A given set of six consecutive hourly cycles provides 17 perturbed members, with the Random Parameter 2 (RP2) stochastic physics scheme employed and IC perturbations supplied by the parent ensemble MOGREPS-G, and one control member (member 0) with unperturbed initial

© Crown copyright 2025, Met Office

conditions and the RP2 scheme turned off.

As the WMV ensemble takes its ICs and LBCs from MOGREPS-UK, it is also a time-lagged ensemble. To generate a daily 18-member ensemble, the 1800 UTC MOGREPS-UK cycle is used, this comprises the three members from each of the 1800, 1700, 1600, 1500, 1400 and 1300 UTC model cycles. The RP2 scheme is not used in the WMV ensemble, members are purely downscaled from MOGREPS-UK.

The WMV was run using the RAL3.1 science configuration whereas MOGREPS-UK still uses RAL2-M. More details about RAL3.1 and RAL2-M are provided in section 3.3.. Both models use the MORUSES urban scheme (Met Office–Reading Urban Surface Exchange Scheme; Porson et al., 2010) but the WMV land-use is derived from the ~300m resolution ESA Climate Change Initiative version 2 (CCIv2) land cover dataset, whereas MOGREPS-UK land-use is derived from the 1990 ~25m resolution Institute of Terrestrial Ecology (ITE) land cover dataset. Although lower resolution than ITE, the CCI dataset is newer so includes more recent urban developments (Figure 3). Soil moisture and MURK aerosol in the WMV are both downscaled from MOGREPS-UK.

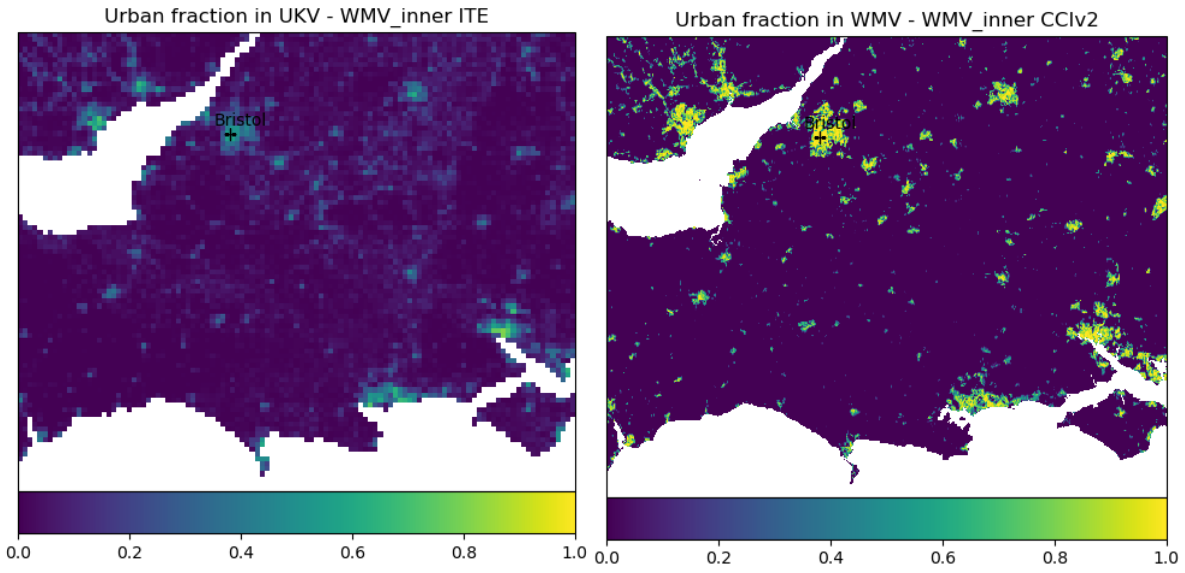


Figure 3. Urban fraction from the UKV which uses ITE as the source dataset (left) and the WMV which uses CCIv2 as the source dataset (right).

3.2. Higher resolution ensembles

Post-event, some of the WesCon Intensive Observation Periods (IOPs) were re-run using the RAL3.1 science configuration in MOGREPS-UK. Along with the WMV,

© Crown copyright 2025, Met Office

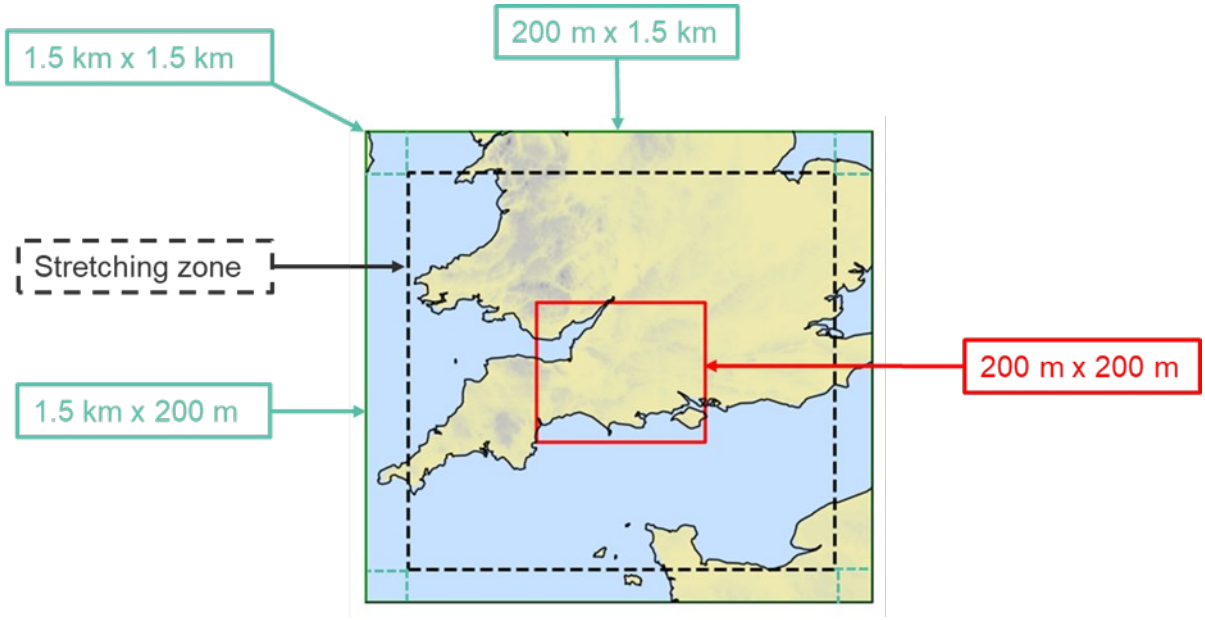
two further variable resolution ensembles with inner grid lengths of 200 m (WMV200) and 100 m (WMV100) were nested inside the RAL3.1 MOGREPS-UK. The inner region of both the WMV200 and WMV100 domains covers the same area as the WMV inner region except with inner grid lengths of 200 m and 100 m respectively (Figure 4). Like the WMV, an inflation factor of 1% was used to increase the grid length gradually to 1.5 km at the edges. Table 2 describes the main differences between the models.

Model	Gridlength (km)	Total grid points	Inner grid points	Levels	Timestep (s)	Science config
MOGREPS-UK	2.2 -> 4.4	740 x 752	421 x 546	70	100	RAL3.1
WMV	0.333 -> 1.5	856 x 774	510 x 421	70	12	RAL3.1
WMV200	0.222 -> 1.5	1204 x 1064	766 x 631	120	6*	RAL3.1
WMV100	0.111 -> 1.5	2094 x 1819	1531 x 1261	120	4	RAL3.1

Table 2. A comparison of the models that will be rerun for all IOPs. *The WMV200 uses a 6s timestep rather than 8s so that STASH diagnostics are available with a frequency of 5 minutes.

The models were run using RAL3.1 as 18-member time-lagged ensembles with 3 members initialised at 13, 14, 15, 16, 17 and 18 UTC and run out for 38 hours. Along with the WMV, these higher resolution domains take their initial and boundary data from MOGREPS-UK, which is rerun using the RAL3.1 science configuration using the archived operational restart files. All models use the MORUSES urban scheme, however while the WMV models use the CCIv2 land use dataset MOGREPS-UK is still using ITE as it is run using the operational ancillary files. Soil moisture and MURK aerosol in the WMV ensembles are both downscaled from MOGREPS-UK.

© Crown copyright 2025, Met Office



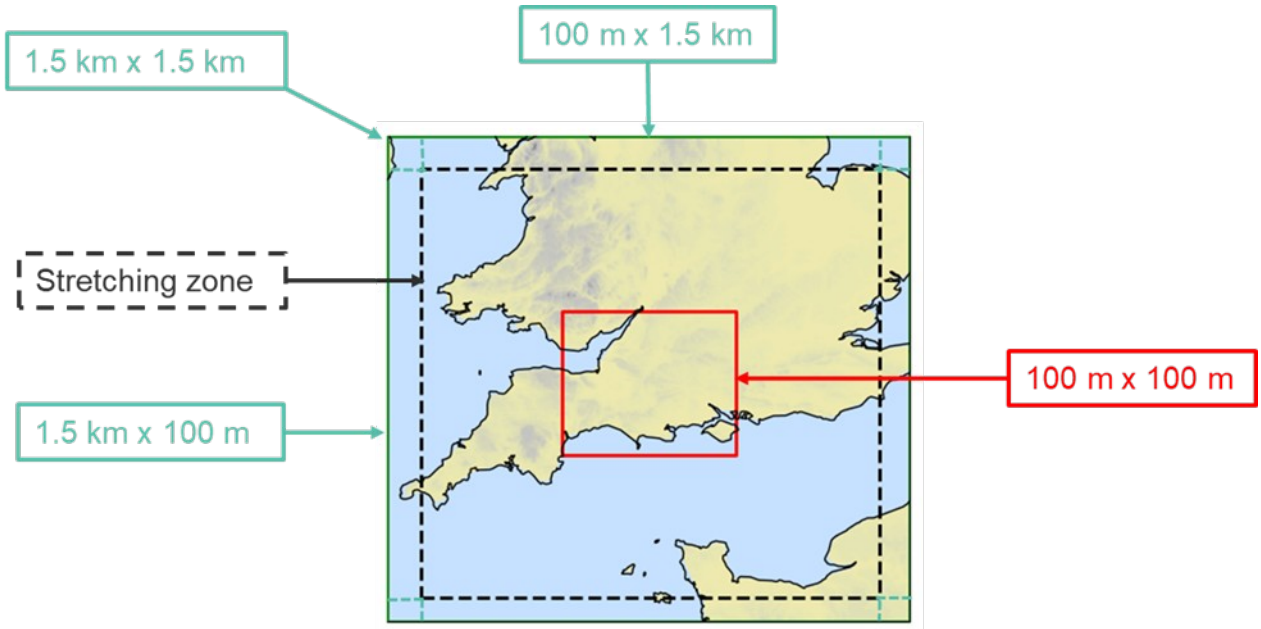


Figure 4. The WMV200 (top) and WMV100 (bottom) domains.

The following cases from summer 2023 have been run to date:

Case	Initialised	Weather Summary
11 June	13-18Z 10 June	A band of Thunderstorms moved through the WesCon region during late afternoon/evening

© Crown copyright 2025, Met Office

15 June	13-18Z 14 June	Cumulus clouds forming along sea breeze
21 June IOP 5	13-18Z 20 June	Showers in the WesCon region during the day
22 June IOP 6	13-18Z 21 June	A few isolated, heavy showers developed in the WesCon region during the afternoon
5 July IOP 8	13-18Z 4 July	Post-frontal showers throughout the day
6 July IOP 9	13-18Z 5 July	A few showers during the morning, mostly dry during the afternoon/evening. SW winds.
11 July IOP 11	13-18Z 10 July	Cold front clearing overnight, intense band of slow-moving precipitation during afternoon.
12 July IOP 12	13-18Z 11 July	Showers associated with trough/occluded fronts
13 July IOP 13	13-18Z 12 July	Showers associated with trough
17 July IOP 14	13-18Z 16 July	Showers associated with trough
19 July	13-18Z 18 July	Occluded fronts clearing overnight, NW-ly flow with a few isolated showers in the WesCon region during the afternoon
25 July IOP 16	13-18Z 24 July	NW-ly flow with a few showers developing in the WesCon region
3 August IOP 18	13-18Z 2 August	Showers developing in NW-ly flow

Table 3. Cases run with RAL3.1 MOGREPS-UK, WMV, WMV200 and WMV100 Ensemble Nesting Suite.

3.3. Science configurations

Since Autumn 2019, the operational UK models (MOGREPS-UK and the UKV) and the routinely running 300m London model have been using the RAL version 2.0 mid-latitude configuration (RAL2-M: Bush et al., 2023). The WesCon experiments detailed above will be run using RAL version 3.1. A full description of RAL3 and in-depth evaluation is provided by Bush et al. (2024). The main differences between RAL2-M and RAL3.1 are shown in Table 4 and described briefly below.

Configuration	Cloud schem	Microphysics scheme	Stochastic PBL	Free atmospheri	Fountain buster

© Crown copyright 2025, Met Office

	e		perturbations	c mixing length	
RAL2-M	Smith	Wilson-Ballard	Yes	Option 2	No
RAL3.1	Bi-modal	CASIM	No	Option 3	Yes

Table 4. Key differences between the science configurations used.

The first difference is the cloud scheme, with RAL2-M using the diagnostic "Smith" large-scale cloud scheme (Smith, 1990) and RAL3.1 using the new bi-modal cloud scheme (Van Weverberg et al., 2021). The bimodal cloud scheme defines entrainment zones around inversions. Within these entrainment zones, two modes of variability are used to calculate the cloud liquid water content and the cloud fraction, conserving the grid-box mean value of saturation departure. The scheme uses the approach of Furtado et al. 2016 to calculate variances of the individual modes of variability (so no pre-specified critical relative humidity profile is required).

The second difference is the microphysics scheme with RAL2-M using the single moment scheme of Wilson and Ballard (1999) and RAL3.1 using the CASIM multi-moment scheme (Field et al., 2023), which permits the UM to have single or double moments microphysical capability. In RAL3.1 it is used as a double moment scheme. As a consequence of using the CASIM microphysics scheme in RAL3.1, the coefficients used in the diagnostic lightning scheme have been retuned. The UM uses the lightning parameterisation of McCaul et al. (2009), which produces a flash rate F as:

$$F = 0.95\,F1 + 0.05\,F2$$

Where $F1 = k_1\left[w q_g(-15\ ^\circ C)\right]$, and $F2 = k_2\left[\int \left[\rho(q_g + q_s + q_i)\right] dz\right]$. Here, $w q_g$ represents the flux of graupel (vertical velocity multiplied by graupel mixing ratio) at the −15°C level in the storm and $\int \left[\rho(q_g + q_s + q_i)\right] dz$ is the total column ice water path (integral of air density multiplied by the sum of graupel, snow, and ice mixing ratios). Previously, the coefficients k_1 and k_2 were taken as in McCaul et al. (2009):

$$k_1 = 0.042 \text{ and } k_2 = 0.2.$$

In RAL3.1, the coefficients used are:

$$k_1 = 0.21 \text{ and } k_2 = 0.60.$$

A third difference between the configurations is the use of stochastic boundary layer perturbations which are applied to the potential temperature and moisture

© Crown copyright 2025, Met Office

fields in the RAL2-M configuration. These were originally included in RAL1-M to help kick-start convection, which was often delayed in kilometre-scale versions of the UM run without a convection scheme in the mid-latitudes. The perturbations were kept in RAL2-M despite the tendency for relatively earlier triggering of convection.

There are also differences in the PBL mixing options. To determine the free atmospheric mixing length used by the 1D PBL scheme, RAL2-M uses an asymptotic (i.e. away from the surface) mixing length of 40m (option 2), whereas RAL3.1 diagnoses turbulent layer depths through the whole atmospheric column and uses those layer depths to determine the mixing length as L=MAX[0.15*layer_depth, 40m] (option 3). So, option 3 can only systematically increase the mixing length (when it finds a “turbulent layer”), and hence is more diffusive. In all three configurations the 1D PBL mixing length is blended with the Smagorinsky subgrid turbulence scheme mixing length (0.2 x the horizontal grid length), using a blending weight that tells us how poorly resolved the turbulence is. In RAL2-M the blending weight is reduced to zero (to give 3D Smagorinsky) with an exponential function of the ratio of height to grid size above the PBL top. RAL3.1 uses a revised blending weight that tends to 1 (the 1D PBL mixing length) outside of turbulent layers. However, within turbulent layers, the revised blending still uses the standard blending formula based on L/Dx, so the proportion of 1D PBL to Smagorinsky depends on how well-resolved the layer depth, L, is compared to the grid size, Dx.

RAL3.1 also includes changes to address the non-conservative behaviour of Semi-Lagrangian advection, including the consistent treatment of potential temperature and moisture variables and the introduction of the “fountain buster” scheme. The fountain buster is a modification to the UM's SL advection scheme aimed at suppressing “eternal fountains” whereby single grid-column updrafts can become unrealistically intense and persistent because a stagnation point forms at the base of the updraft. The scheme works by identifying grid-points where the winds are converging into a stagnation point and applies a simple linear up-wind advection increment to add in the effects of convergent in-flow missed by the SL advection.

Finally, there are several changes introduced at RAL3.1 that attempt to consolidate regional and global land configurations (see Bush et al., 2024 for details).

3.4. Verification metrics

A range of different verification techniques have been used to assess the

© Crown copyright 2025, Met Office

characteristics of the simulated precipitation forecasts against observations. The observations used here come from the Met Office radar rainfall composite (Harrison *et al.*, 2011). The radar data used to generate the composite consists of 5-min scan sequences at four elevations from the 15 C-band radars across the UK, at a resolution of 600 m in range and 1° in azimuth. The final, quality-controlled radar rainfall composite is produced on a $1 \times 1\,\mathrm{km}^2$ grid. Here, all data (both model and observations) are first aggregated to the coarsest model grid for a fairer comparison.

Probabilities of exceeding a given rainfall threshold can be produced by averaging over the raw model data. For high-resolution forecasts, where predictability is low at the grid scale (Lorenz, 1969), the probability fields can be noisy, and signals of extremes can be weak. One way to avoid this issue is to use a neighbourhood processing method to smooth the probabilities (e.g., Schwartz and Sobash, 2017) where information from the surrounding grid points is used to calculate the probabilities. Schwartz and Sobash describe two different neighbourhood approaches to produce probabilities that can be interpreted as either (i) a likelihood of event occurrence at the grid scale or (ii) a likelihood of event occurrence over spatial scales larger than the grid scale. The first addresses the uncertainty at the grid scale due to insufficient ensemble members, while the second can amplify weak signals of extremes. Here the second approach, known as the neighbourhood maximum probability, is adopted. In this method a neighbourhood length x is chosen; at each grid point all neighbourhood within x km are searched to determine whether the event occurred, with the grid point assigned a value of 1 if the event occurred anywhere in the neighbourhood and 0 if not. The neighbourhood search is performed for each ensemble member and a probability that an event occurred within a vicinity of x km is then obtained by averaging across all members. Here the area in the vicinity of each grid point is defined using a square neighbourhood of length 8 MOGREPS-UK grid boxes ($\sim 17.5 \times 17.5\,\mathrm{km}^2$), noting that the WMV data are first aggregated to the MOGREPS-UK grid. Using this method, the ensemble probability is then calculated, providing a ‘worst case scenario’ for the precipitation events.

The first verification technique used is convective “cell statistics” (Hanley et al., 2015), an object-based statistical technique for assessing the size and intensity of convective storms in both the simulated and observed precipitation fields. A “storm” is identified as a contiguous area of rainfall exceeding a given rainfall threshold (usually 1 or 4 mm/hr) and covering a minimum area of 4 grid boxes. Distributions of storm equivalent diameter and intensity can then be constructed, with the equivalent diameter of a storm defined as the diameter of a circle with the same area.

The second verification technique used is the Fractions Skill Score (FSS: Roberts

© Crown copyright 2025, Met Office

and Lean, 2008). The FSS is a neighbourhood-based verification metric used to compare two precipitation fields (e.g., model and observations). It is computed by first converting each grid point of the precipitation fields to binary, depending on whether they exceed a chosen precipitation threshold. Then for each grid square, comparing how many points in each field exceed the threshold in the surrounding area as defined by a “neighbourhood size” (e.g., a neighbourhood size of 3 refers to the nine grid squares centred on the square of interest).

Three FSS metrics are used here, the FSS metrics for assessing ensemble spatial spread and skill (Dey et al., 2014; 2016a) and the FSS metric for spatial displacement (Skok and Roberts, 2017). The spatial dispersion (dFSS) metric compares the spatial agreement of every member-member pair in the ensemble for a range of different neighbourhood sizes. The domain-averaged value of the dFSS is calculated for every lead time and every neighbourhood. The values of dFSS range from 0 to 1 with low (high) values representing high (low) spread. An equivalent measure for the spatial skill (or error) can be found by computing the average of the FSS for every member-obs pair. This value, termed the eFSS, also ranges from 0 to 1 with high (low) values representing small (large) errors. Dey et al. (2014) take a dFSS or eFSS value of approximately 0.5 as the scale at which the ensemble is seen to be believable or ‘skillful’.

Roberts and Lean (2008) and Skok (2015, 2016) showed that the FSS is related to the spatial displacement between the observed and simulated precipitation fields. Skok and Roberts (2017) provide a “recipe” for using the FSS to compute the displacement between two fields, which is summarised below:

1. Threshold the observed and simulated precipitation fields using either a percentile or an absolute threshold. An absolute threshold should only be used if the frequency bias between the two fields is less than 2.
2. Calculate the FSS at the grid scale (FSS_{n1}) i.e., there is no neighbourhood only the binary values.
3. Reassign the non-zero pixels that overlap from both fields to zero, so you are only left with the pixels in each field that are not coincident with the other field.
4. Find the smallest neighbourhood length n that gives you an FSS value above 0.5, FSS_n.
5. Compute the displacement for the non-overlapping fields as:

$$d_{FSSno} = \left(1 - FSS_n\right) \bullet n$$

Since FSS_n is close to 0.5, this can be approximated to:

$$d_{FSSno} = n/2$$

6. The mean displacement between the binary pixels in the original two

© Crown copyright 2025, Met Office

fields, D_{FSS}, can then be computed as:

$$D_{FSS}=\left(1-FSS_{n1}\right)\bullet d_{FSSno}$$

Here the dFSS, eFSS and D_{FSS} have been calculated for the WMV and MOGREPS-UK ensembles, using the radar composite as observations. To ensure the neighbourhood lengths and domain size are equivalent for both ensembles, the WMV data has been regridded to the 2.2km MOGREPS-UK grid and the metrics have been calculated over the WMV domain for both ensembles. The metrics have been applied to the 95th and 99th percentile of hourly rainfall accumulations. Using a percentile threshold avoids a bias in the number of points compared in each member-member and member-observation combination, allowing the spatial distribution of rainfall to be evaluated.

4. Routine running results

4.1. Comparison of the WMV with MOGREPS-UK

In this section the RAL3.1 WMV ensemble has been compared to the output from the operational RAL2-M 1800 UTC MOGREPS-UK cycle. A selection of cases has been chosen to highlight some of the strengths and weaknesses of the WMV. Additional case studies are presented in the Summer 2023 UK Testbed Sub-km Convection Characteristics Report (Flynn et al., 2023).

4.1.1. Case 1 (IOP 8): Post-frontal showers aided by surface convergence

On 5 July 2023 post-frontal showers developed in a westerly flow aided by low-level surface convergence. Figure 5 shows the domain-average precipitation for this day from the WMV ensemble compared with the radar-derived surface rainfall rate composite. Most WMV members had the onset time of precipitation between 2-3 hours too early, with the peak precipitation occurring ~3 hours too early and being overestimated by as much as 3x by some members. The afternoon/evening precipitation (15-20 UTC) and precipitation decay were better represented by most members, with the ensemble mean closely matching the radar observations.

© Crown copyright 2025, Met Office

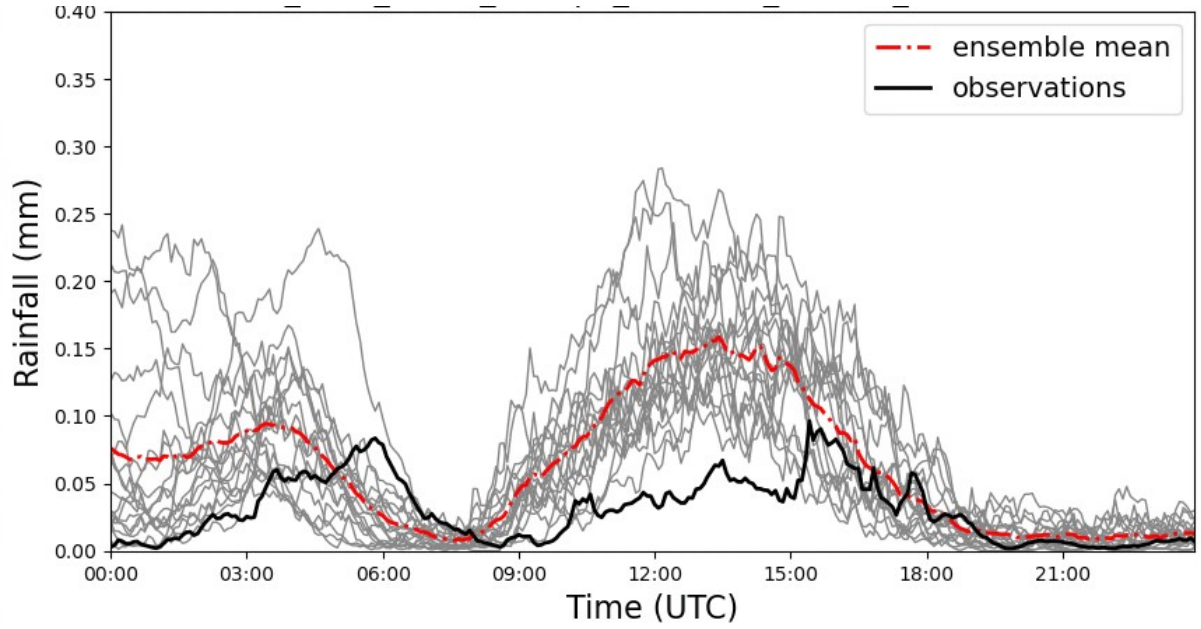


Figure 5. Domain-average precipitation for 5 July 2023 from the WMV ensemble members (grey lines) and the radar composite (black line). The ensemble mean is indicated by the red line. Data is averaged over the WMV inner region.

Figure 6 shows the surface rainfall rate postage stamps from the WMV ensemble and MOGREPS-UK at 14 UTC compared with the radar-derived surface rainfall rate composite. The WMV ensemble has a lot of very small, scattered showers, but is better able to capture the organisation of the precipitation along the convergence line than MOGREPS-UK. This leads to more reliable probabilities for heavy precipitation (Figure 7).

© Crown copyright 2025, Met Office

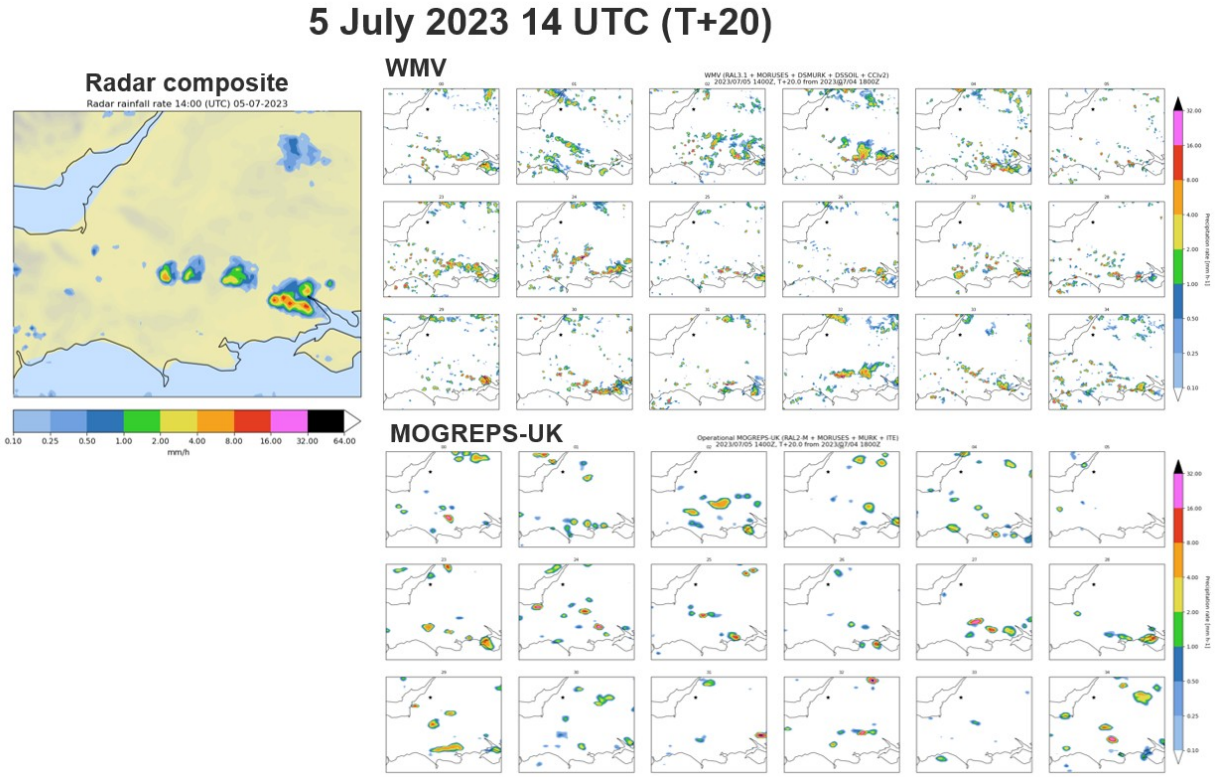


Figure 6. Surface rainfall rate at 1400 UTC 5 July 2023 from the WMV ensemble (top) and the operational 1800 UTC MOGREPS-UK ensemble (bottom) compared with the radar composite over the inner WMV domain.

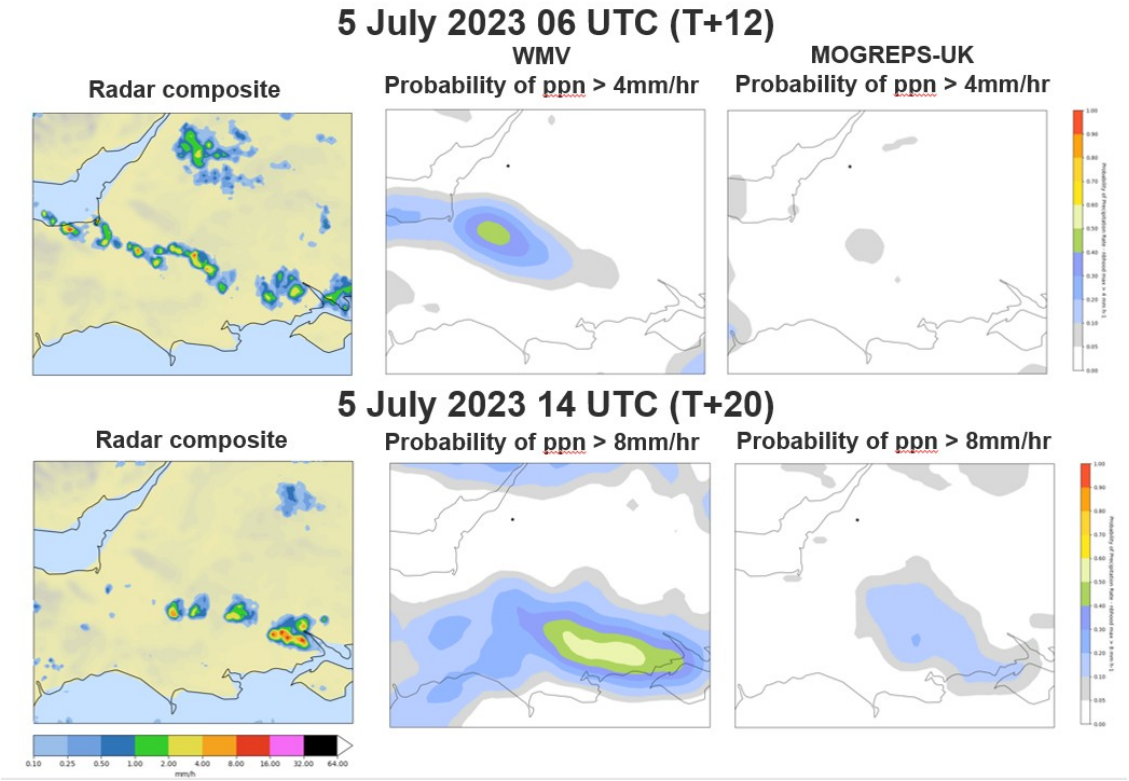


Figure 7. Radar-derived surface rainfall rate at 0600 UTC (top) and 1400 UTC (bottom) 5 July 2023 over the inner WMV domain. Also shown are the neighbourhood maximum ensemble probability plots of rainfall rate exceeding 4 mm/hr (top) and 8 mm/hr (bottom) in the vicinity of each grid

© Crown copyright 2025, Met Office

point for a neighbourhood of 17.5 x 17.5 km^2 from MOGREPS-UK (right) and the WMV ensemble (middle).

The probabilities from the WMV ensemble give better guidance than MOGREPS-UK with high probabilities for precipitation in the correct location. However, it's noted that the WMV ensemble has too high probabilities in the north of the domain at 14 UTC.

Figure 8 shows the average storm size against average storm mean rainfall rate for each member of the MOGREPS-UK and WMV ensembles compared with the radar observations for 5 July 2023.

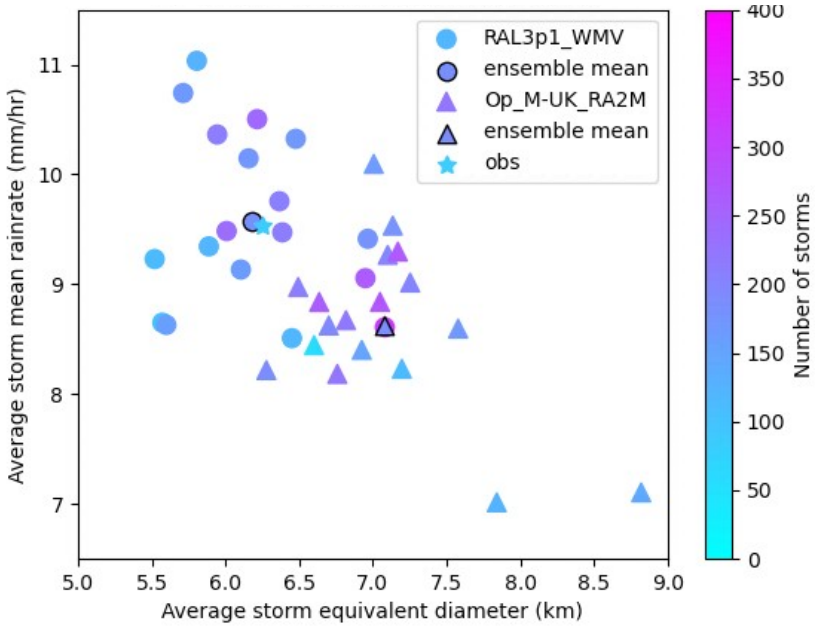


Figure 8. Average storm equivalent diameter against average storm mean rainfall rate for the period 9-21 UTC 5 July 2023 for each member of the MOGREPS-UK (diamonds) and WMV (circles) ensembles. The ensemble mean is indicated by the black outline. Data has been regridded to 2.2 km and extracted over the inner WMV domain. Colours indicate the number of storms identified and the radar observations are indicated by the star. Storms are identified using a rainfall rate threshold of 4 mm/hr and a minimum area of 4 MOGREPS-UK grid boxes (19 km^2). Data is every 5 minutes.

Most MOGREPS-UK members had too few storms which were on average larger and weaker in intensity than the observed storms. In contrast the WMV members tend to have too many small, intense storms although on average the size and intensity of the WMV storms compares better to the radar-observed storms.

4.1.2. Case 2 (IOP 11): Organised deep convection

On 11 July 2023 a cold front cleared to the south of England late morning which

© Crown copyright 2025, Met Office

introduced a very unstable WNW'ly flow and lead to a convergence line of organised showers forming across the Southwest into Central England that afternoon. Figure 9 shows the surface rainfall rate postage stamps from the WMV ensemble and MOGREPS-UK at 15 UTC compared with the radar-derived surface rainfall rate composite.

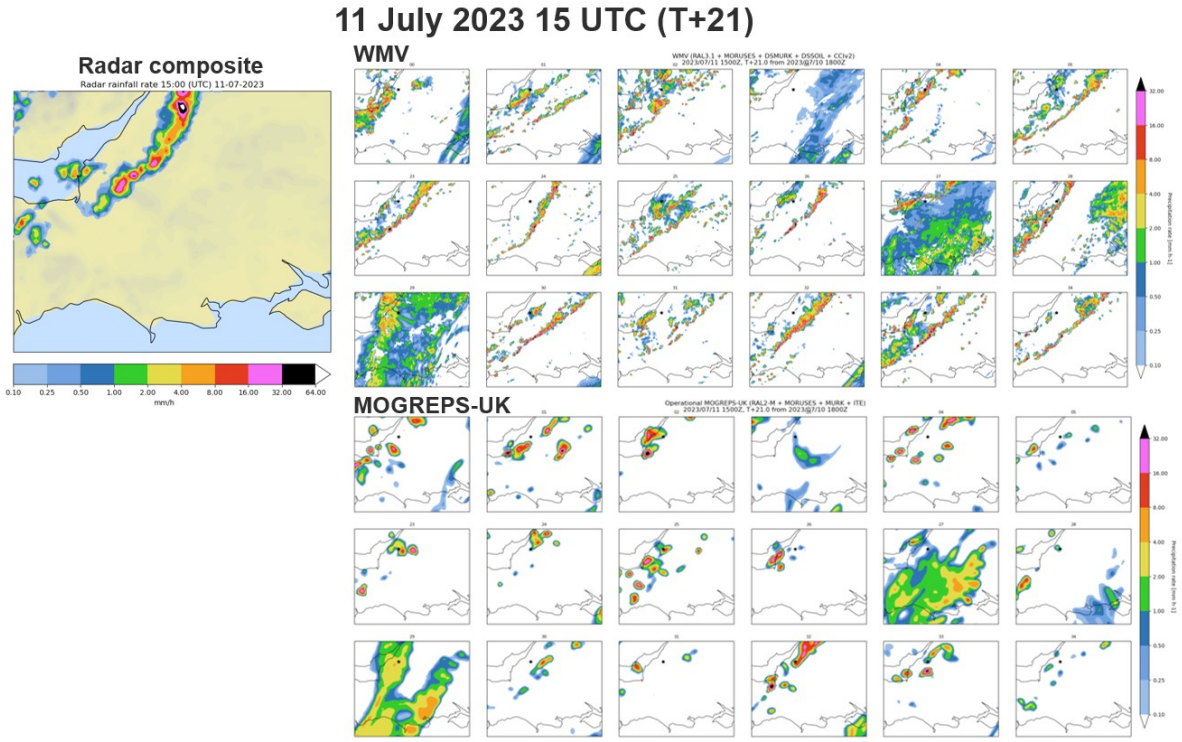


Figure 9. Surface rainfall rate at 1500 UTC 11 July 2023 from the WMV ensemble (top) and the operational 1800 UTC MOGREPS-UK ensemble (bottom) compared with the radar composite over the inner WMV domain.

The position and the mode of the convection is inherited from the driving ensemble e.g. a couple of members of the WMV ensemble are simulating widespread heavy precipitation and so are the corresponding driving MOGREPS-UK members. The biggest difference between the two ensembles is the organisation of the convection, with about half the WMV members simulating an organised band of precipitation whereas MOGREPS-UK misses the organisation of the band and has mostly isolated, blobby cells. As a result, the WMV has more extensive and greater probabilities of precipitation exceeding 8 mm/hr, the WMV probabilities also have stronger indication of there being a band of heavy precipitation (Figure 10).

While the WMV had a good representation of the organised band of precipitation, the postage stamps in Figure 9 highlight that the WMV members tended to simulate more small showers around the band than were observed by the radar. This is quantified in Figure 11, which shows the average storm size against average storm mean rainfall rate for each member of the MOGREPS-UK and LMV

© Crown copyright 2025, Met Office

ensembles compared with the radar observations for 11 July 2023.

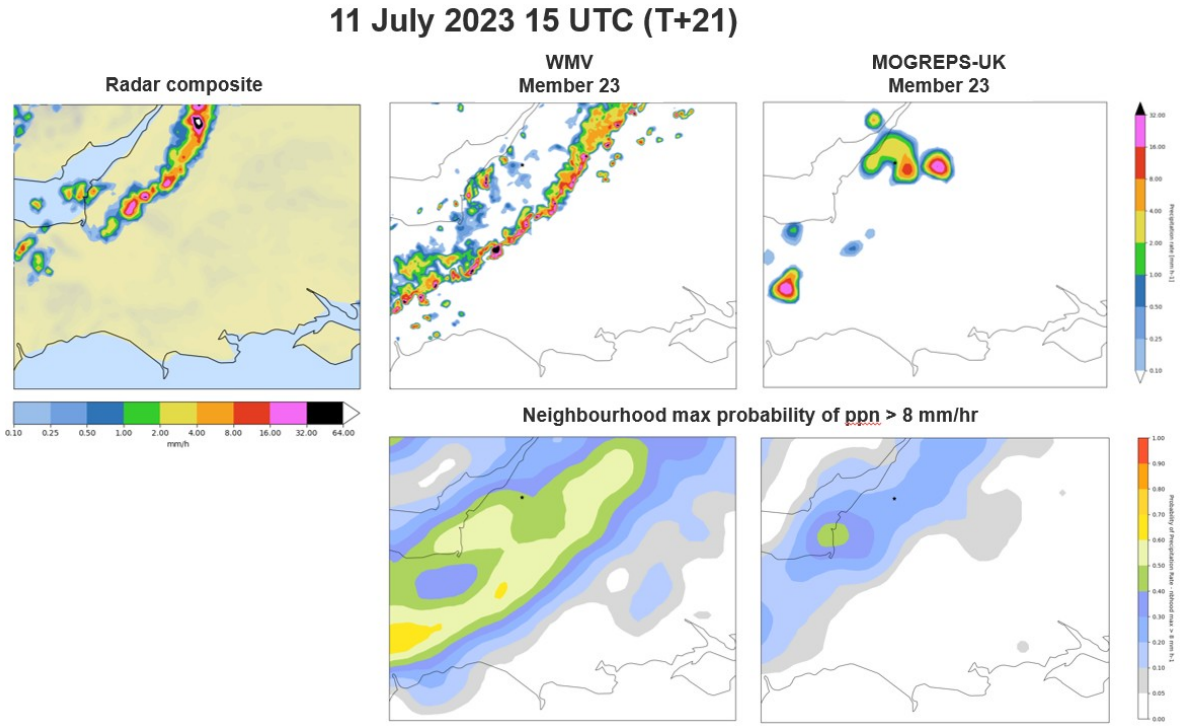


Figure 10. Surface rainfall rate at 1500 UTC 11 July 2023 from member 23 of both the WMV ensemble and the operational 1800 UTC MOGREPS-UK ensemble compared with the radar composite over the inner WMV domain. Also shown are the neighbourhood maximum ensemble probability plots of rainfall rate exceeding 8 mm/hr in the vicinity of each grid point for a neighbourhood of 17.5 x 17.5 km^2 from MOGREPS-UK and the WMV ensemble.

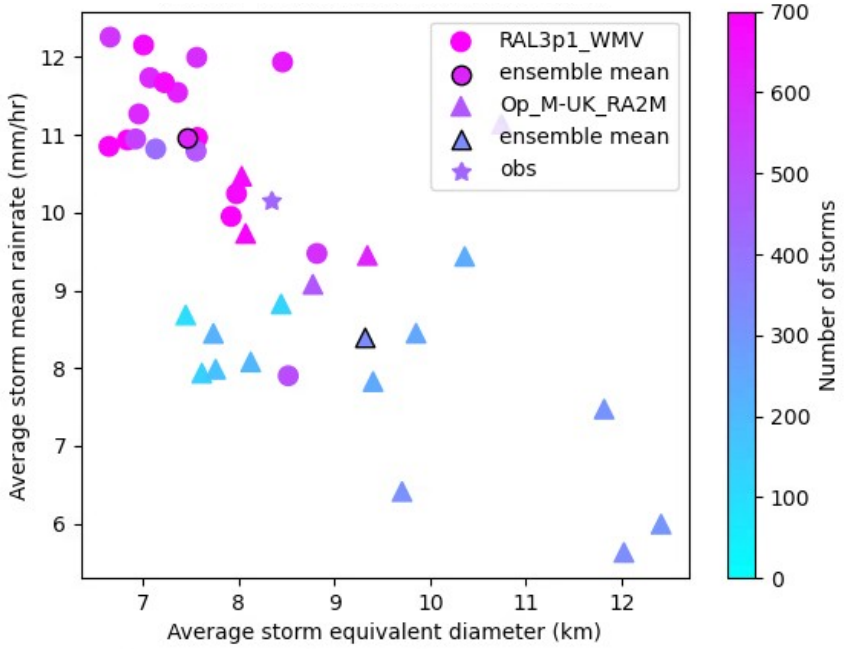


Figure 11. Average storm equivalent diameter against average storm mean rainfall rate for the

© Crown copyright 2025, Met Office

period 9-21 UTC 11 July 2023 for each member of the MOGREPS-UK (diamonds) and WMV (circles) ensembles. The ensemble mean is indicated by the black outline. Data has been regridded to 2.2 km and extracted over the inner WMV domain. Colours indicate the number of storms identified and the radar observations are indicated by the star. Storms are identified using a rainfall rate threshold of 4 mm/hr and a minimum area of 4 MOGREPS-UK grid boxes (19 km^2). Data is every 5 minutes.

The WMV members tend to simulate too many storms which are on average smaller and more intense than the observed storms. In contrast, most MOGREPS-UK members simulate too few storms that are on average larger and less intense than the observed storms.

During the Summer 2023 UK Testbed it was noted that on some occasions there are very high 10m wind gusts associated with the heavy precipitation in the WMV (Glazer et al., 2023). An example of this is shown in Figure 12, which shows the scale-dependent 10m wind gust (max in hour) between 14-15 UTC on 11 July 2023 from member 32 of the WMV and MOGREPS-UK compared with the SYNOP station observations of maximum gust speed. Also shown is the probability of the scale-dependent 10m wind gust exceeding 40 kts.

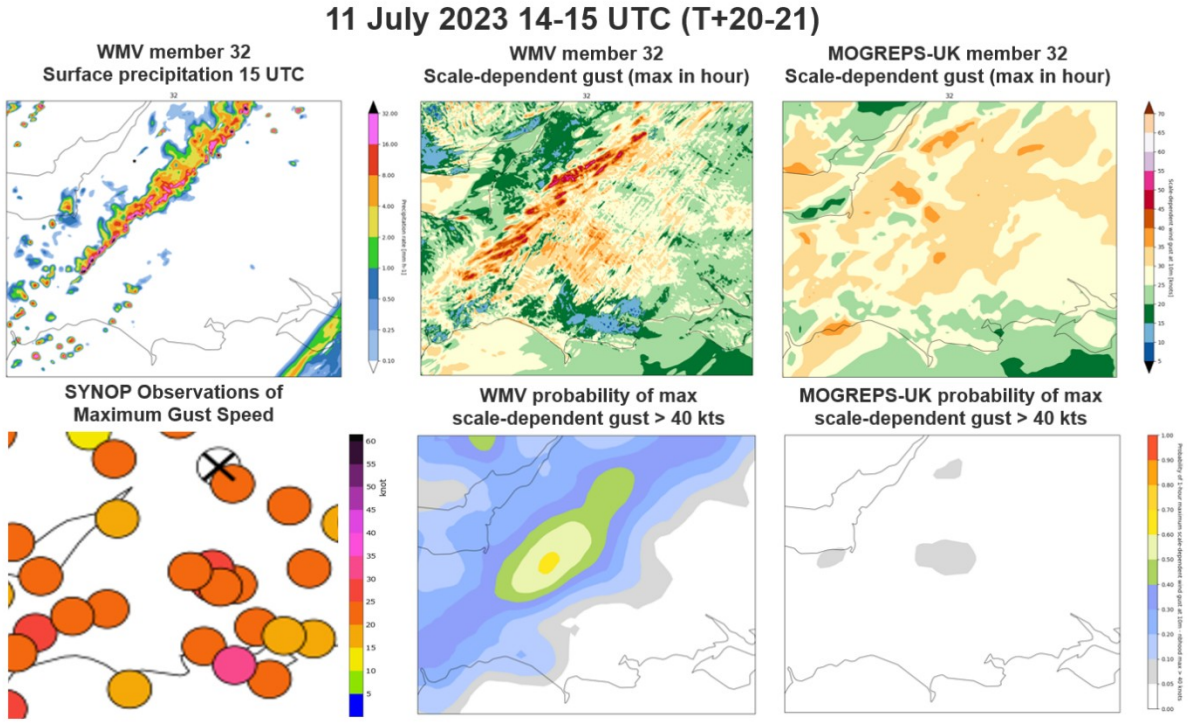


Figure 12. Scale-dependent 10m wind gust (max in hour) for 1400-1500 UTC 11 July 2023 from member 32 of both the WMV ensemble and the operational 1800 UTC MOGREPS-UK ensemble compared with the SYNOP station observations for the maximum gust speed that hour over the inner WMV domain. Also shown are the neighbourhood maximum ensemble probability plots of the maximum scale-dependent 10m wind gust exceeding 40 kts in the vicinity of each grid point for a neighbourhood of 17.5 x 17.5 km^2 from MOGREPS-UK and the WMV ensemble.

© Crown copyright 2025, Met Office

The WMV has strong 10m wind gusts associated with the heavy precipitation, with gusts widely greater than 40 kts and in a few locations exceeding 50 kts. This leads to widespread probabilities of 10m wind gusts exceeding 40 kts. In contrast, the 10m wind gusts associated with the precipitation in MOGREPS-UK are generally between 30-40 kts, with only a very small probability for gusts exceeding 40 kts. While it's expected to have strong gusts associated with a squall line, the WMV gusts may be too strong. The SYNOP station observations widely have gusts between 25-30 kts in the vicinity of the precipitation band, however there is only one station that recorded a maximum gust exceeding 30 kts, and that station (Portland Bill) was far ahead of the convergence line. However, it should be noted that between 14-15 UTC the squall line was situated in a region with a lack of SYNOP stations so the maximum wind gusts may have been missed.

4.1.3. Case 3 (non-IOP): Diurnal shallow convection

On 19 July 2023 a few diurnally-forced showers developed in the WesCon region in a NW-ly flow. A strong capping-inversion at ~750 hPa resulted in mostly shallow convection. Figure 13 shows the surface rainfall rate postage stamps from the WMV ensemble and MOGREPS-UK at 12 UTC compared with the radar-derived surface rainfall rate composite.

© Crown copyright 2025, Met Office

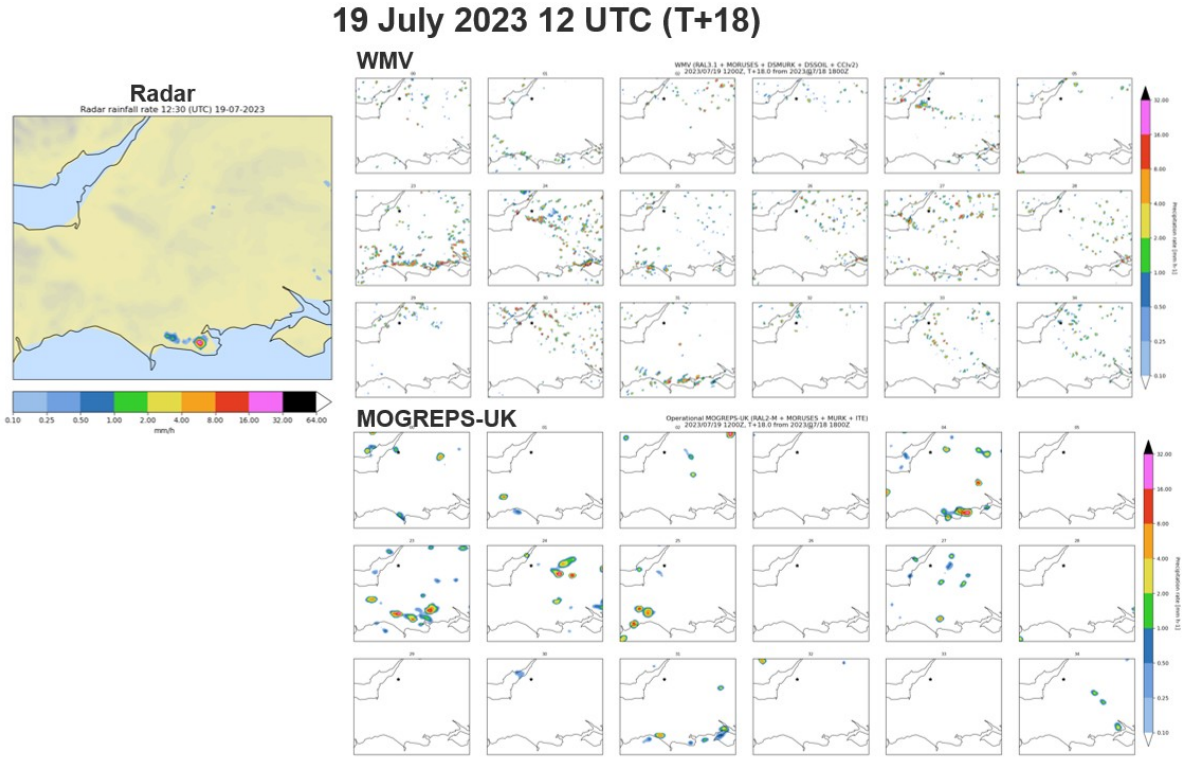


Figure 13. Surface rainfall rate at 1200 UTC 19 July 2023 from the WMV ensemble (top) and the operational 1800 UTC MOGREPS-UK ensemble (bottom) compared with the radar composite over the inner WMV domain.

Most members of the WMV ensemble have too many small, intense, scattered showers. MOGREPS-UK provides better guidance, having fewer, more isolated storms. Comparing both the observed and simulated surface rainfall with the cloud top height, precipitating showers generally have cloud top heights exceeding 9 kft/3 km (Figure 14). The WMV, which has too many small precipitating showers, has more clouds with tops exceeding 3 km than were observed. Comparing a member with a lot of showers (4) to a member with fewer showers (5), the member with fewer precipitating showers has fewer clouds with tops exceeding 3 km. This suggests the WMV has too many small precipitating showers due to shallow clouds getting too deep.

© Crown copyright 2025, Met Office

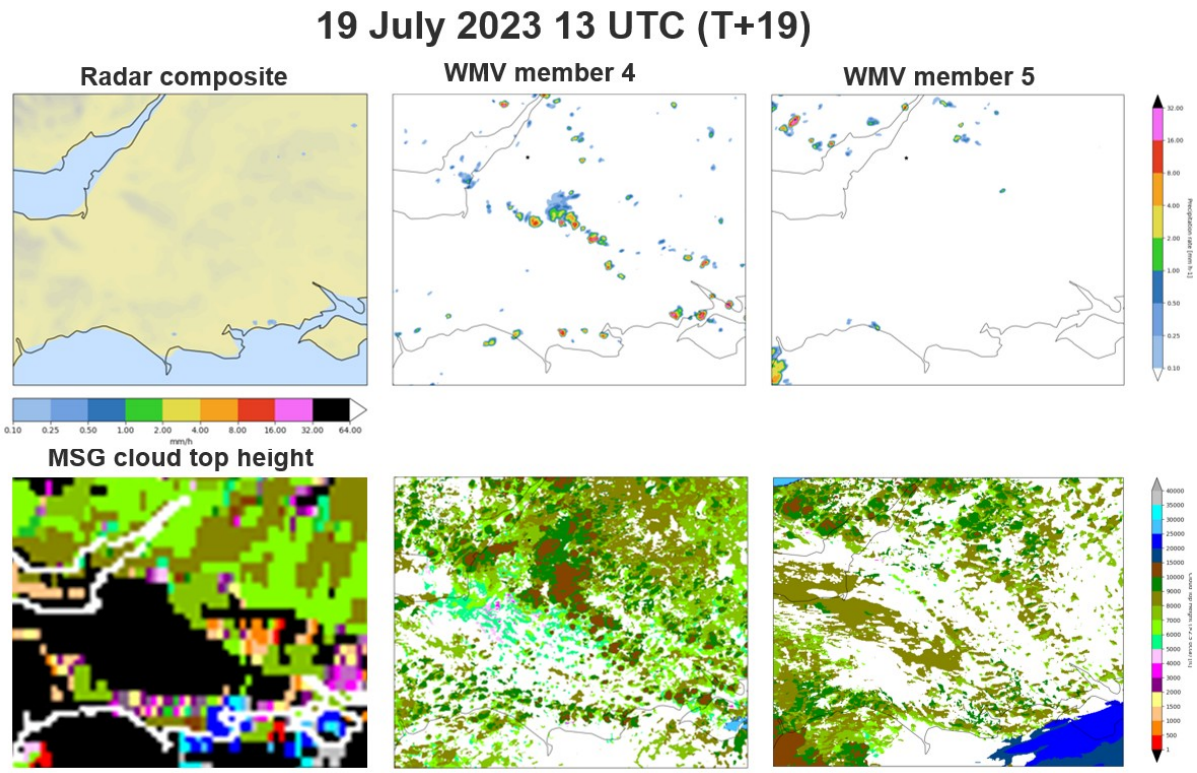


Figure 14. Surface rainfall rate from members 4 and 5 of the WMV ensemble compared with the radar composite (top) and cloud top height from the same WMV members compared with the MSG satellite cloud top height product (bottom) at 1300 UTC 19 July 2023 over the inner WMV domain.

4.2. FSS Verification

The FSS metrics for the WMV and operational MOGREPS-UK have been calculated for 14 June - 25 August 2023 (excluding 30 July – 6 August due to missing radar and model data) using the 95th and 99th percentile of hourly rainfall accumulation. Figure 15 shows the distribution of hourly rainfall accumulation for the radar observations and the unperturbed member 0 of the WMV and MOGREPS-UK ensembles for the period 14 June – 25 August 2023. Also shown is the distribution of the 95th and 99th percentile hourly rainfall accumulation for the same period.

© Crown copyright 2025, Met Office

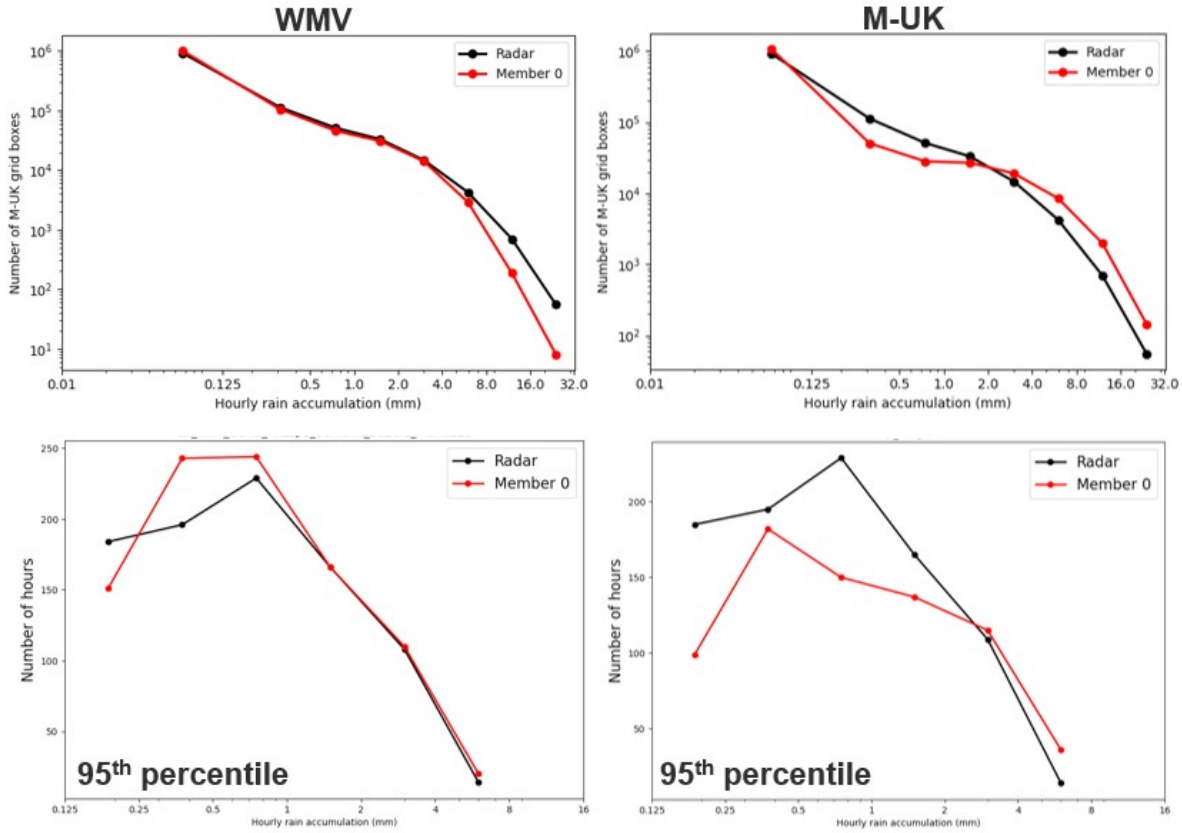


Figure 15. Distribution of hourly rainfall accumulation for the radar observations and the unperturbed member 0 of the WMV (left) and MOGREPS-UK (right) ensembles for the period 14 June – 25 August 2023 (excluding 30 July – 6 August). Also shown is the distribution of the 95th percentile of hourly rainfall accumulation for the same period.

The WMV has a good representation of light to moderate hourly rainfall accumulation events (less than 4 mm) but under-represents heavier rainfall events (hourly accumulations greater than 8 mm). While the individual storms in the WMV have been shown to have too intense instantaneous rainfall rates (e.g. Figure 11), they also tend to be too small which could result in lower hourly rainfall accumulations. Additionally, the WMV storms could be too short-lived, resulting in lower rainfall accumulations. In contrast MOGREPS-UK under-represents light rainfall events (less than 2 mm) and over-represents heavy rainfall events (greater than 4 mm). This is again counter to the individual storm statistics which show MOGREPS-UK storms tend to have less intense instantaneous rainfall rates compared to the observed storms (e.g. Figure 11). They do however also tend to be too large which may lead to overestimating the hourly rainfall totals. In addition, they could be too long-lived.

The distribution of the 95th percentile hourly rainfall accumulation has a peak between 0.5-1 mm for the radar observations. For the WMV the peak is shifted slightly lower (with roughly the same number of occurrences in the 0.25-0.5 mm and 0.5-1 mm bins) and for MOGREPS-UK it is between 0.25-0.5 mm. Despite

© Crown copyright 2025, Met Office

differences in the shape of the distribution, the mean value (averaged over all cases and lead times) for the 95th percentile hourly rainfall accumulation is ~0.5 mm for both the WMV and MOGREPS-UK ensembles and the radar observations. The 99th percentile corresponds to an average hourly rainfall accumulation of ~1.5 mm.

Figure 16 shows the eFSS and dFSS calculated for the 95th percentile of hourly rainfall accumulation. Similar results are obtained using the 99th percentile (not shown).

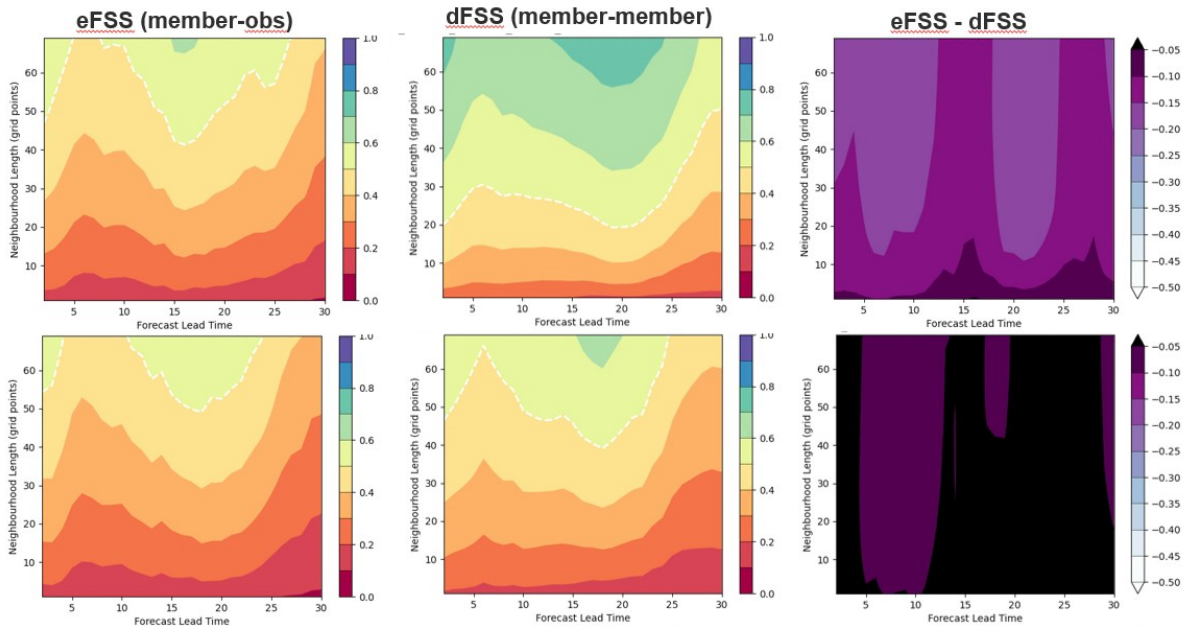


Figure 16. Contour plots showing the dFSS, eFSS and their difference (eFSS minus dFSS) for the WMV ensemble (top) and MOGREPS-UK (bottom). Metrics are calculated using the 95th percentile of hourly rainfall accumulation averaged over 14 June – 25 August 2023 (excluding 30 July – 6 August). All data has been regridded to 2.2km and extracted over the WMV domain.

For both ensembles, the dFSS and eFSS contours generally slope upwards indicating that the spatial spread and error both grow from small to medium scales with increasing lead time. However, both ensembles have a peak in dFSS and eFSS at lead times of ~T+6, suggesting larger spatial spread and error during the night than during the day (for reference T+0 is 18 UTC). Even at very short lead times, the eFSS values are less than 0.5 at small scales (less than 30 km) suggesting there is very little skill at these scales. At most lead times, the neighbourhood length scale at which the eFSS exceeds 0.5 is slightly lower for the WMV ensemble compared to MOGREPS-UK suggesting that the WMV is more skilful at smaller scales than MOGREPS-UK. While the skilful scales for the eFSS are similar for the WMV ensemble and MOGREPS-UK, the dFSS values are larger for MOGREPS-UK than the WMV ensemble at all neighbourhood lengths and forecast lead times indicating there is less spatial spread in the WMV ensemble than in MOGREPS-UK. For a perfect ensemble, the spatial spread should capture

© Crown copyright 2025, Met Office

the spatial error so that the dFSS and eFSS scales match meaning the ensemble has the correct spatial spread-skill relationship (e.g., Dey et al., 2016b). The difference between dFSS and eFSS gives an indication of whether the ensemble is over- or under-spread. For MOGREPS-UK, the difference is small showing it has a good spatial spread-skill relationship. However, for the WMV ensemble the difference between eFSS and dFSS is slightly negative suggesting it is under-spread. Similar results were obtained for the 300m variable resolution London model (LMV) ensemble run during summer 2022 (Hanley and Lean, 2024). MOGREPS-UK RAL3 trials are also slightly under-spread (Anne McCabe, personal communication) suggesting it is the different physics rather than the increased horizontal resolution leading to the difference in spatial spread. It is hypothesised that the increase in light precipitation associated with the double-moment CASIM microphysics scheme may be leading to a reduction in spatial spread.

Figure 17 shows the displacement FSS, D_{FSS}, for each member-obs pair of the WMV ensemble vs the member-obs pair of the equivalent driving MOGREPS-UK member. The D_{FSS} values have been averaged over forecast lead times T+2-T+30 (where T+0 is 18 UTC) across the summer 2023 trial period.

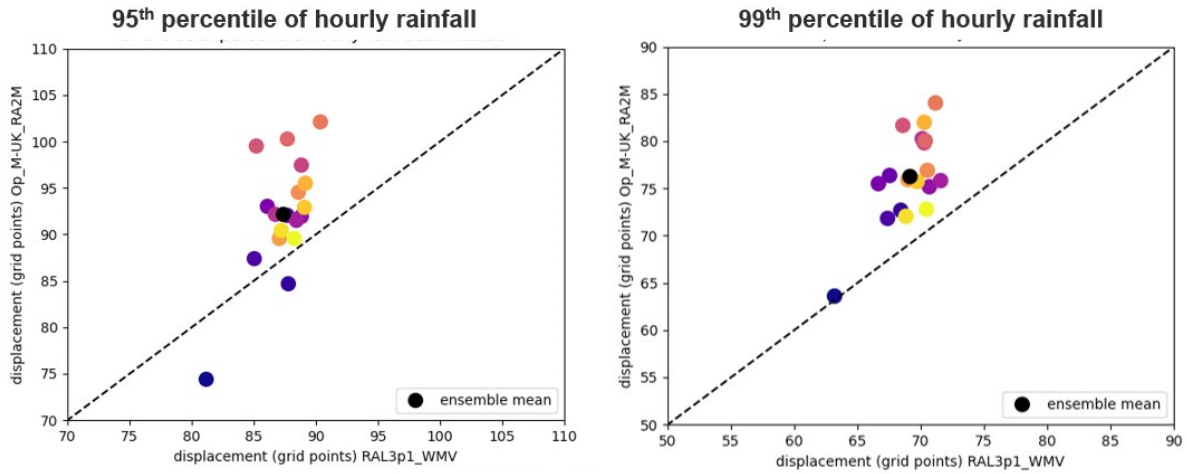


Figure 17. Displacement FSS for each member of the WMV ensemble (x-axis) against the equivalent driving MOGREPS-UK member (y-axis). FSS metrics are calculated using the 95th percentile of hourly rainfall accumulation averaged over T+2-30 for the period 14 June – 25 August 2023 (excluding 30 July – 6 August). All data has been regridded to 2.2km and extracted over the WMV domain.

In general, most values lie above the 1-1 line showing that D_{FSS} is less in the WMV than in the driving MOGREPS-UK member. This indicates that the displacement between the simulated and observed precipitation is less in the WMV ensemble than in MOGREPS-UK.

© Crown copyright 2025, Met Office

5. Post-campaign ensemble simulations

In this section two case studies are presented from the ensemble simulations that have been completed post-campaign (as described in Section 3.2)

5.1. IOP 5: Diurnal convection associated with low-level convergence

On 21 June 2023 diurnally-forced convective storms developed along a low-level convergence line. Figure 18 shows the domain-average precipitation for this day from the MOGREPS-UK, WMV and WMV100 ensembles compared with the radar-derived surface rainfall composite.

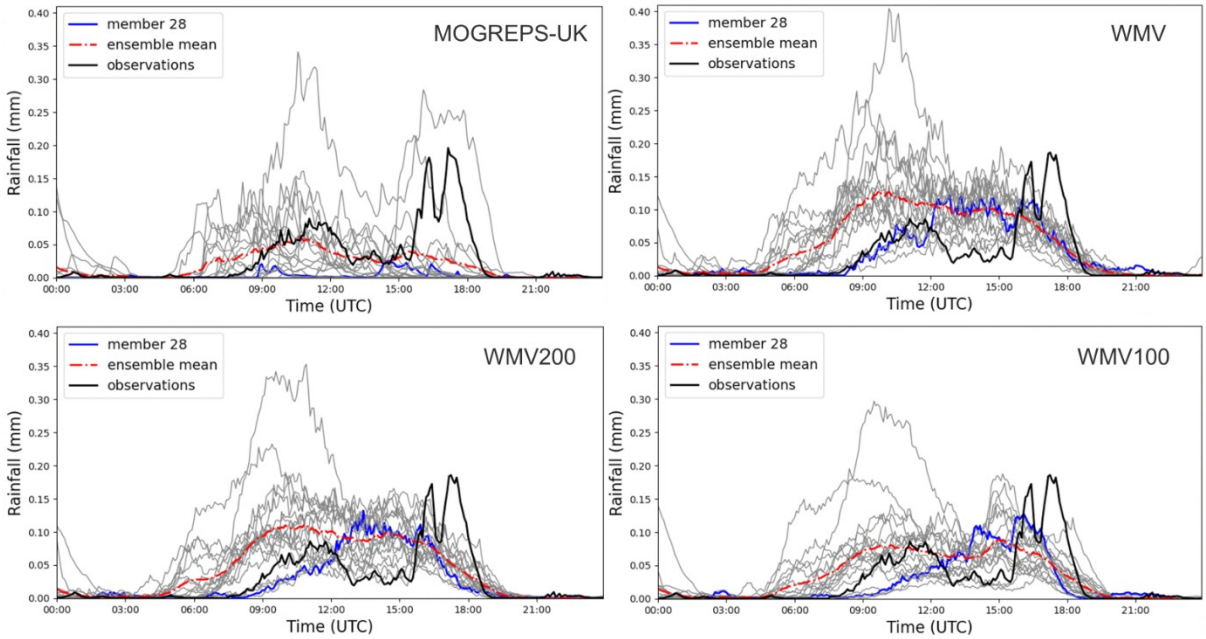


Figure 18. Domain-average precipitation for 21 June 2023 from the MOGREPS-UK, WMV, WMV200 and WMV100 ensembles compared with the radar composite (black line). Individual members are indicated by the grey lines and the ensemble mean is indicated by the red line. Data is averaged over the WMV inner region.

Most members of all four ensembles have a too early onset time of precipitation, with the WMV being earliest. The radar observations showed two peaks in the precipitation, the first at ~11 UTC with a second larger peak in the afternoon. The ensembles all captured the morning peak; however, in the WMV and WMV200 ensembles it was on average ~ 2 hours too early, and the magnitude was overestimated by the ensemble mean. The timing of the peak was better captured by MOGREPS-UK and the WMV100 ensemble and while the MOGREPS-

© Crown copyright 2025, Met Office

UK ensemble mean underestimated the magnitude of the peak, the WMV100 ensemble mean simulated a peak of the correct magnitude. The afternoon peak was also too early in all four ensembles, occurring at ~15 UTC in the ensemble mean, with the magnitude of the peak underestimated by each ensemble mean. The WMV100 was the only ensemble that correctly simulated the 2nd peak as being larger than the first, although the magnitude of the 2nd peak was underestimated by the ensemble mean. Most MOGREPS-UK members have precipitation dying out after 15 UTC, although a couple of members were able to simulate the afternoon peak.

Figure 19 shows the surface rainfall rate from member 28 (indicated by the blue line in Figure 18) of the RAL3.1 MOGREPS-UK, WMV, WMV200 and WMV100 ensembles at 12 UTC compared with the radar-derived surface rainfall rate composite. This member was chosen for having the “best” onset time of precipitation in the WMV.

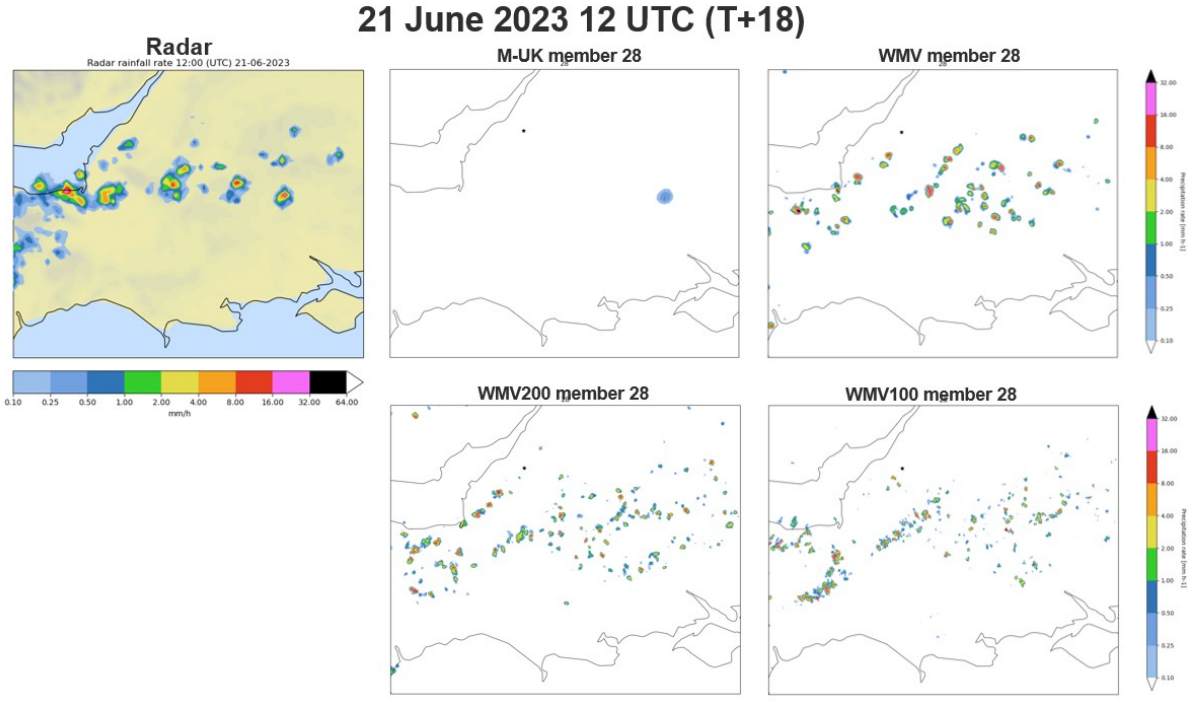


Figure 19. Surface rainfall rate at 1200 UTC 21 June 2023 from member 28 of the RAL3.1 MOGREPS-UK, WMV, WMV200 and WMV100 ensembles compared with the radar composite over the inner WMV domain.

A line of convective showers associated with low-level convergence extended most of the way across the inner WMV domain. Several members of MOGREPS-UK missed the showers or only had one or two light showers (e.g. member 28). All members of the WMV, WMV200 and WMV100 ensembles simulated showers in the region, however they are too keen to organise the showers into NE-SW orientated lines. Most members of the WMV have too many small, intense

© Crown copyright 2025, Met Office

showers. As the resolution is increased, the showers get even smaller with the WMV100 having the smallest showers, however the shower intensity in the WMV100 appears reduced compared to the WMV and WMV200.

Figure 20 shows the neighbourhood maximum probability of surface rainfall rate exceeding 2 mm/hr for 12 UTC from the MOGREPS-UK, WMV, WMV200 and WMV100 ensembles.

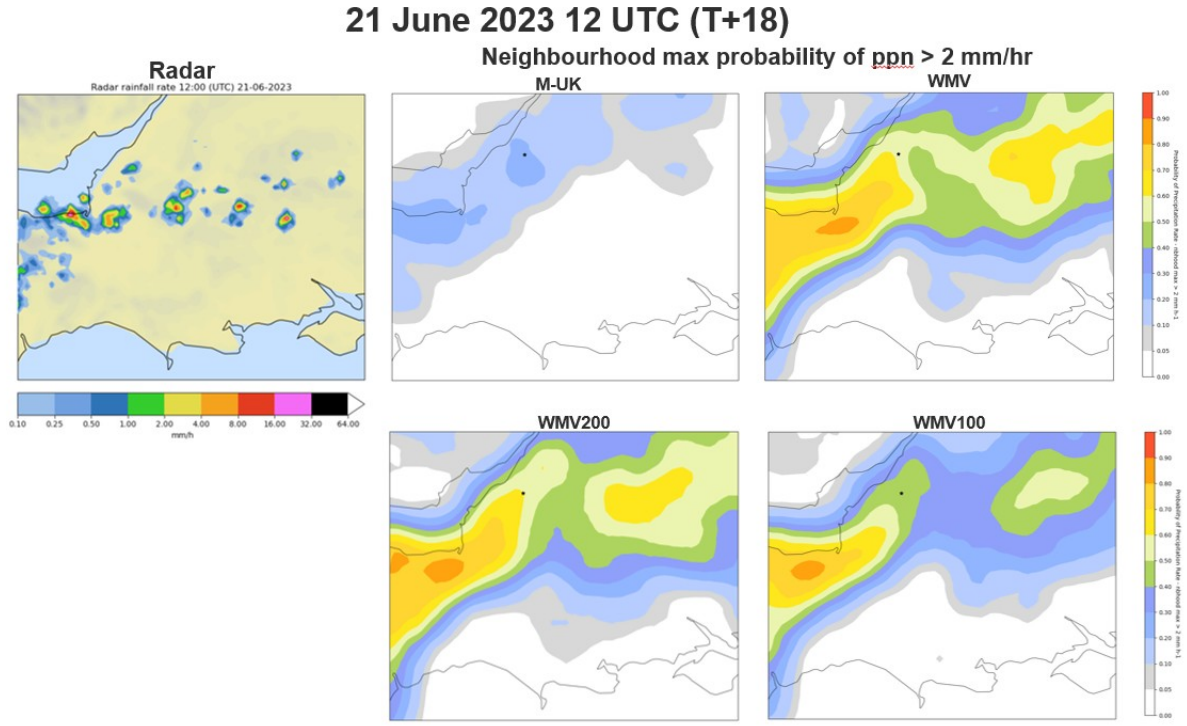


Figure 20. Neighbourhood maximum ensemble probability plots of rainfall rate exceeding 2 mm/hr in the vicinity of each grid point for a neighbourhood of 17.5 x 17.5 km^2 at 1200 UTC 21 June 2023 from the RAL3.1 MOGREPS-UK, WMV, WMV200 and WMV100 ensembles over the inner WMV domain. Also shown is the surface rainfall rate from the radar composite.

The WMV, WMV200 and WMV100 probabilities of surface rainfall rate exceeding 2 mm/hr are more extensive and have larger values than MOGREPS-UK. The MOGREPS-UK probabilities are shifted too far north and don't encompass all the observed storms. The WMV, WMV200 and WMV100 probabilities give a better indication of the location of the showers, although the largest probabilities in the East of the domain are still located too far north. The WMV100 probabilities are slightly reduced compared to the WMV and WMV200 ensembles, consistent with the showers being less intense.

Figure 21 shows the average storm size against average storm mean rainfall rate for each member of the MOGREPS-UK and LMV ensembles compared with the

© Crown copyright 2025, Met Office

radar observations for 11 July 2023.

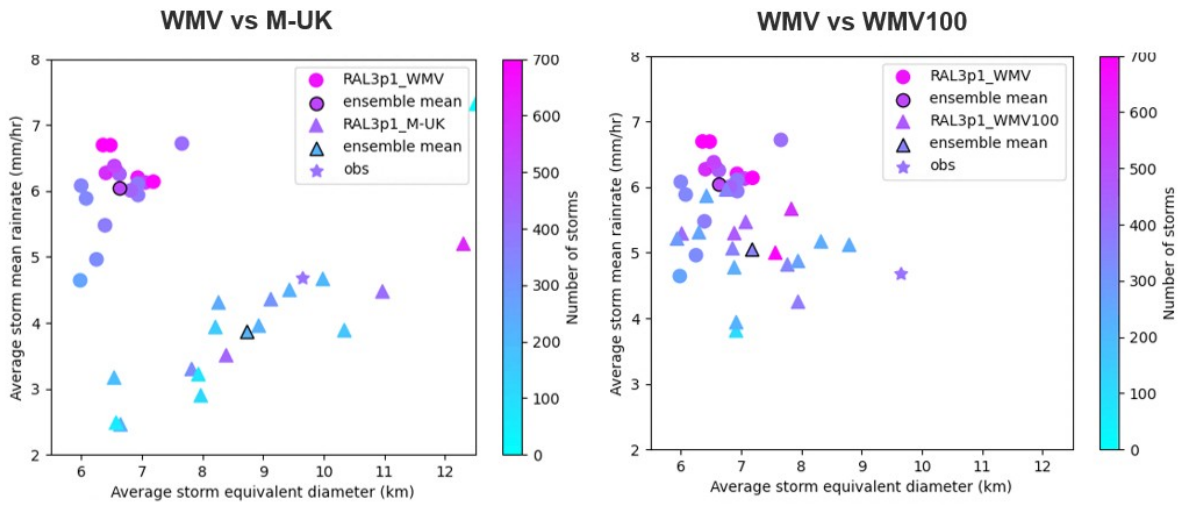


Figure 21. Average storm equivalent diameter against average storm mean rainfall rate for the period 9-21 UTC 21 June 2023 for each member of the WMV (circles) vs MOGREPS-UK (diamonds; left) and WMV100 (diamonds; right). The ensemble mean is indicated by the black outline. Data has been regridded to 2.2 km and extracted over the inner WMV domain. Colours indicate the number of storms identified and the radar observations are indicated by the star. Storms are identified using a rainfall rate threshold of 1 mm/hr and a minimum area of 4 MOGREPS-UK grid boxes (19 km^2). Data is every 5 minutes.

Most members of the WMV ensemble have too many storms, on average the WMV storms are too intense and too small compared to the observed storms. MOGREPS-UK has fewer storms than the WMV, with most members having fewer storms than observed. On average they are less intense than both the observed and the WMV storms, although the average size is closer to the average observed storm size. The WMV100 ensemble has fewer storms than the WMV, with most members having fewer storms than observed. On average they are slightly larger than the WMV storms but smaller than the observed storms and the average intensity (which is less than the WMV) agrees well with the radar observations. It is slightly surprising that the ensemble mean average storm size in the WMV100 ensemble is larger than the WMV ensemble mean average storm size as by-eye the WMV100 appears to have smaller showers (Figure 19). However, many of the smallest showers are likely excluded from the cell statistics as they do not meet the “storm” criteria of having a rainfall rate exceeding 1 mm/hr over a contiguous area covering at least 19 km^2.

On this day, there was a FAAM flight which lasted approximately 5 hours and flew through several of the showers. During this time, the NCAS radars (CAMRa and Kepler) were also scanning, allowing for co-located FAAM and radar measurements of the same clouds. Figure 22 shows a time-series of altitude and vertical velocity from the flight. Take-off was around 0930 UTC, the in-flight

© Crown copyright 2025, Met Office

altitude varied from ~400m to ~4 km with most cloud sampling occurring in the range 2 – 3.2 km. The strongest updraft and downdraft speeds occur in-cloud, with in-cloud values ranging from -8 to 8 m/s and non-cloud values ranging from -4 to 4 m/s.

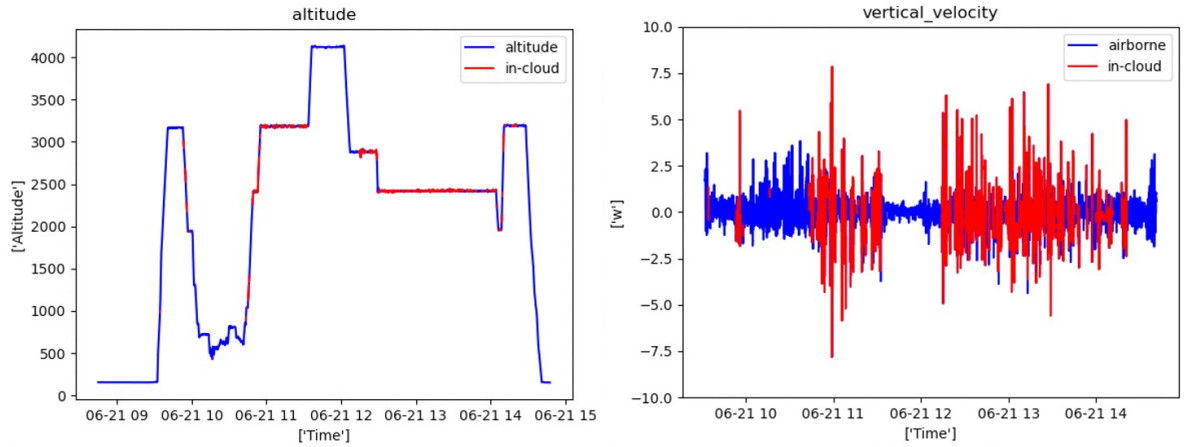


Figure 22. Time-series of altitude (metres) and vertical velocity (m/s) for FAAM flight C342 on 21 June 2023. In-cloud measurements are indicated by the red line.

Figure 23 shows PDFs of updraft velocity (w > 0 m/s) at “all-points” (where the aircraft was flying straight and level) and “in-cloud” (liquid water content > 0.05 g/m^3) from members 27 and 28 of MOGREPS-UK, the WMV and WMV100 compared with the FAAM observations between 10-14 UTC on 21 June 2023. Due to differences in the temporal and spatial sampling between the model and observations, all horizontal grid points in the WMV inner domain have been included (with no regridding of model data) which is why PDFs have been calculated rather than histograms. Based on the flight altitude data (Figure 22), for the PDFs of “all-points” updraft velocity the model data has been constrained in the vertical to be between 0.43 – 4.14 km and for the PDFs of “in-cloud” updraft velocity the model data has been constrained in the vertical to be between 2 - 3.2 km.

© Crown copyright 2025, Met Office

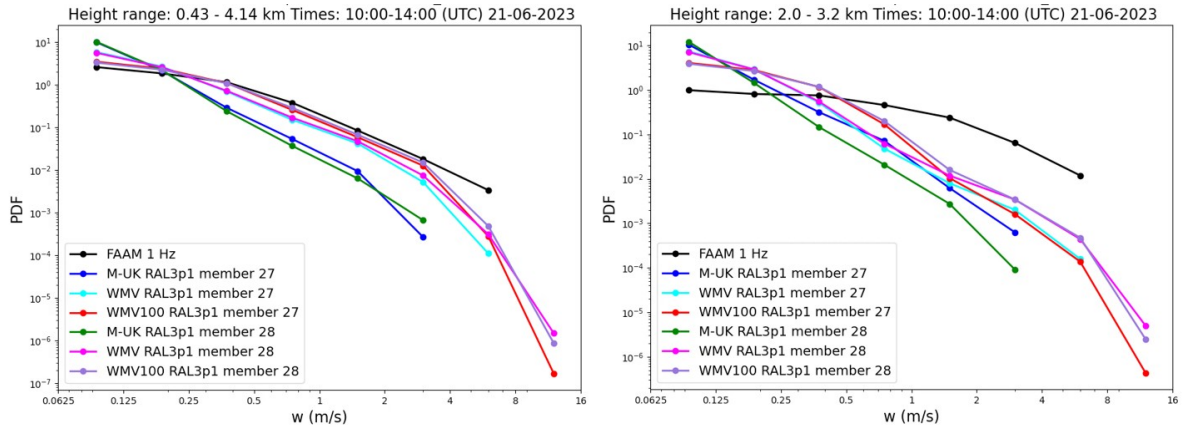


Figure 23. PDF of updraft velocity (w > 0 m/s) at “all-points” (left) and “in-cloud” (right) from members 27 and 28 of MOGREPS-UK, the WMV and WMV100 compared with the FAAM observations on 21 June 2023. Model data is hourly between 10-14 UTC, including all horizontal grid points in the WMV inner domain and all vertical grid points between 0.43 – 4.14 km (left) and 2 – 3.2 km (right). Observations are every second between 0930-1430 UTC.

The WMV and WMV100 PDFs for “all-points” compare well to the FAAM observations, having a slightly larger fraction of updraft velocities less than 0.25 m/s and a slightly smaller fraction of updraft velocities greater than 0.25 m/s. There is a larger difference for the stronger updraft velocities, with the observations having a larger fraction of updraft velocities exceeding 4 m/s than the WMV and WMV100 (noting there are no observed updraft velocities exceeding 8 m/s). The MOGREPS-UK members lie further from the observations, with a larger fraction of updraft velocities less than 0.25 m/s, a smaller fraction with updraft velocities exceeding 0.25 m/s and no updraft velocities exceeding 4 m/s. The “in-cloud” PDFs compare less favourably, with all models over-estimating the fraction of updraft velocities less than 0.25 m/s and underestimating the fraction of updraft velocities exceeding 0.25 m/s. The WMV100 members lie closest to the observations and the MOGREPS-UK members lie furthest from the observations. A possible reason why the “in-cloud” model-obs comparison is worse than the “all-points” is due to the aircraft only sampling a limited number of clouds. By sampling all clouds in the domain, the model PDF may include updrafts from smaller clouds and decaying clouds which have weaker updraft velocities.

It is hard to directly compare the FAAM observations of vertical velocity with the model as during a flight the aircraft is only able to sample a few clouds, at a single level with very high temporal resolution. In contrast, the model has a 3D field of vertical velocity so it can sample all clouds in the domain but due to the large amount of data that entails, the temporal resolution is usually quite low (hourly in this case). One of the novel aspects of WesCon was having co-located observations of vertical velocity from FAAM and the radar retrievals. As part of the ParaChute programme, the radar retrievals of vertical velocity will be verified

© Crown copyright 2025, Met Office

against the FAAM observations which will then hopefully allow the model to be verified against the 3D radar retrievals.

5.2. IOP 6: Diurnal shallow convection

On 22 June 2023 there was mainly non-precipitating shallow cloud with a few diurnally-forced showers developing in the WesCon region during the afternoon. Figure 24 shows the surface rainfall rate postage stamps from the WMV ensemble and MOGREPS-UK at 12 UTC compared with the radar-derived surface rainfall rate composite.

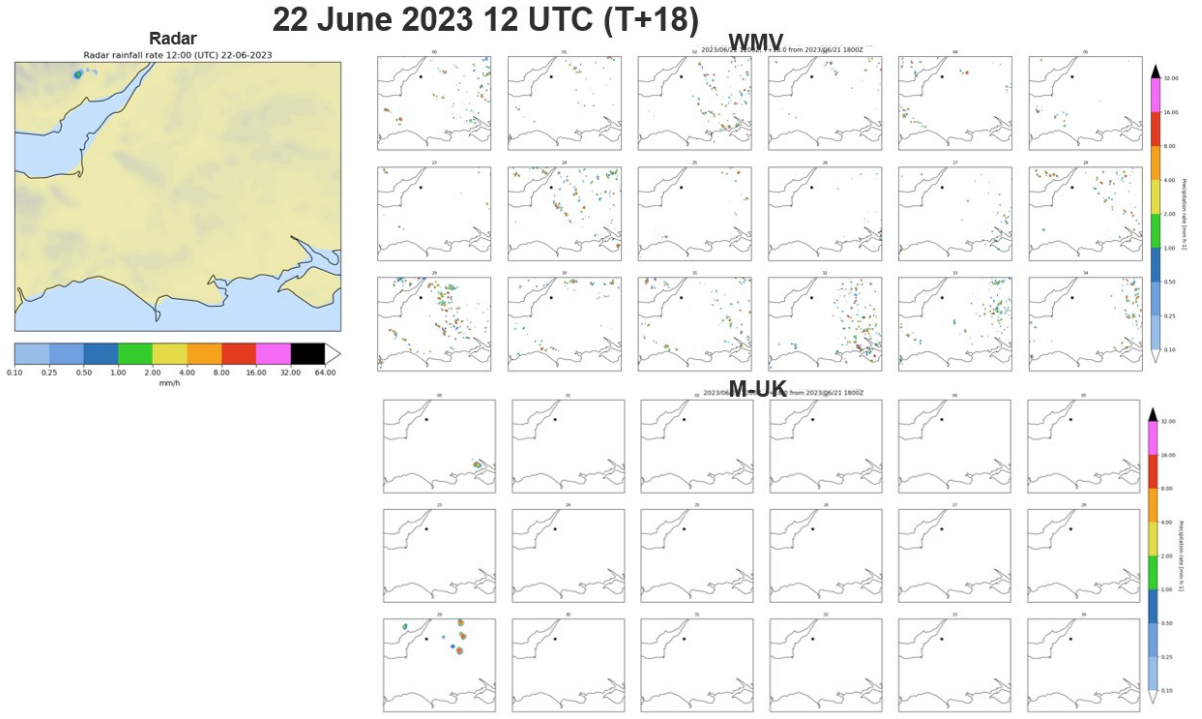


Figure 24. Surface rainfall rate at 1200 UTC 22 June 2023 from the RAL3.1 WMV ensemble (top) and the RAL3.1 MOGREPS-UK ensemble (bottom) compared with the radar composite over the inner WMV domain.

Most members of the WMV ensemble initiated convection too early and had too many widespread, small precipitating showers. This resulted in the WMV over-doing the probability of precipitation exceeding 2 mm/hr (Figure 26). In contrast, most MOGREPS-UK members initiated convection too late, with some members missing the precipitation entirely (e.g. member 23, see Figure 27).

Figure 25 shows the surface rainfall rate postage stamps from the WMV and WMV100 ensembles at 12 UTC compared with the radar-derived surface rainfall rate composite. Some members of the WMV100 ensemble also initiated

© Crown copyright 2025, Met Office

precipitating showers too early. Overall, the WMV100 had fewer, less intense showers than the WMV. This resulted in the WMV100 having reduced probabilities of precipitation exceeding 2 mm/hr compared to the WMV (Figure 26).

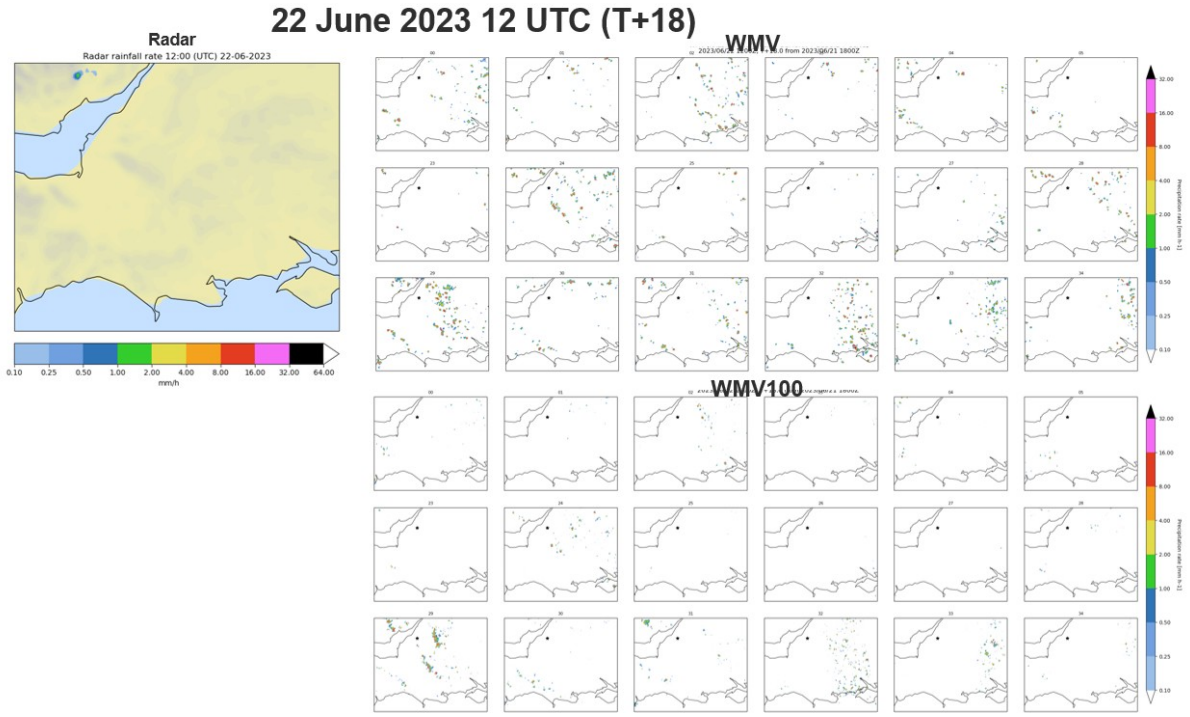


Figure 25. Surface rainfall rate at 1200 UTC 22 June 2023 from the RAL3.1 WMV ensemble (top) and the RAL3.1 WMV100 ensemble (bottom) compared with the radar composite over the inner WMV domain.

© Crown copyright 2025, Met Office

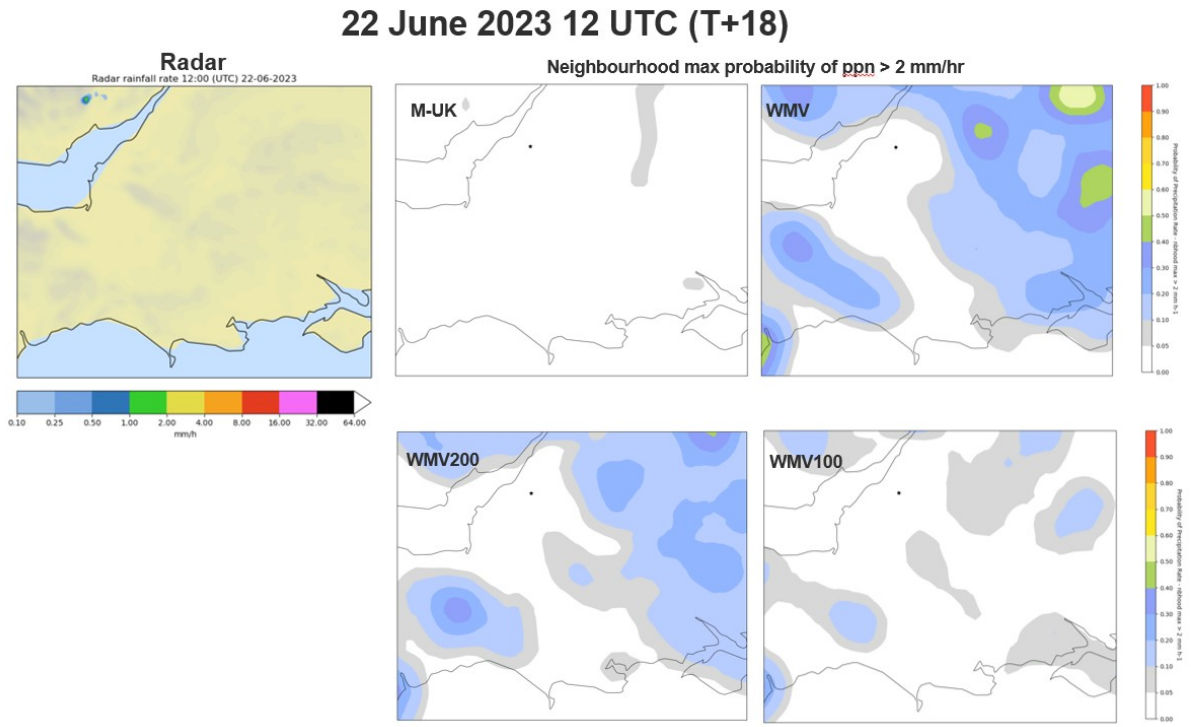


Figure 26. Neighbourhood maximum ensemble probability plots of rainfall rate exceeding 2 mm/hr in the vicinity of each grid point for a neighbourhood of 17.5 x 17.5 km^2 at 1200 UTC 22 June 2023 from the RAL3.1 MOGREPS-UK, WMV, WMV200 and WMV100 ensembles over the inner WMV domain. Also shown is the surface rainfall rate from the radar composite.

Figure 27 shows the domain-averaged precipitation from MOGREPS-UK, the WMV and WMV100 for one member (24) with a lot of precipitating showers compared with one member (23) with fewer precipitating showers.

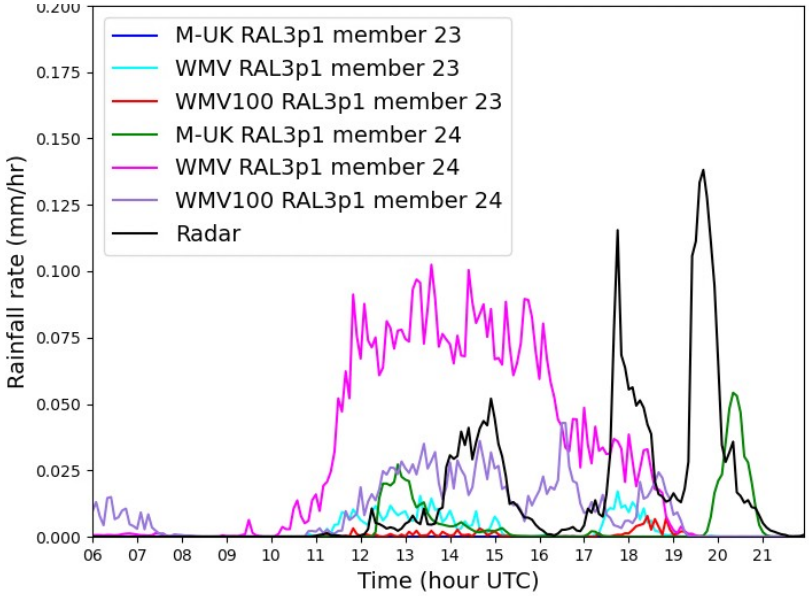


Figure 27. Time-series of domain-average surface rainfall rate for 22 June 2023 from member 23 and member 24 of the RAL3.1 MOGREPS-UK, WMV and WMV100 ensembles compared with the radar composite over the inner WMV domain.

© Crown copyright 2025, Met Office

The onset of precipitation in the WMV100 is ~1 hour later than in the corresponding WMV member and ~1 hour earlier than the corresponding MOGREPS-UK member. In terms of the amount of rainfall, the WMV100 again lies in the middle, having more rainfall than the corresponding MOGREPS-UK member and less rainfall than the corresponding WMV member.

It has been shown previously in this report (e.g. Figure 14) and in previous work (e.g. Hanley and Lean, 2024) that the excessive precipitating showers are usually associated with too deep clouds. This is also true in this case, Figure 28 shows the observed and simulated surface rainfall rate and cloud top height from member 24 of the WMV and MOGREPS-UK at 12 UTC.

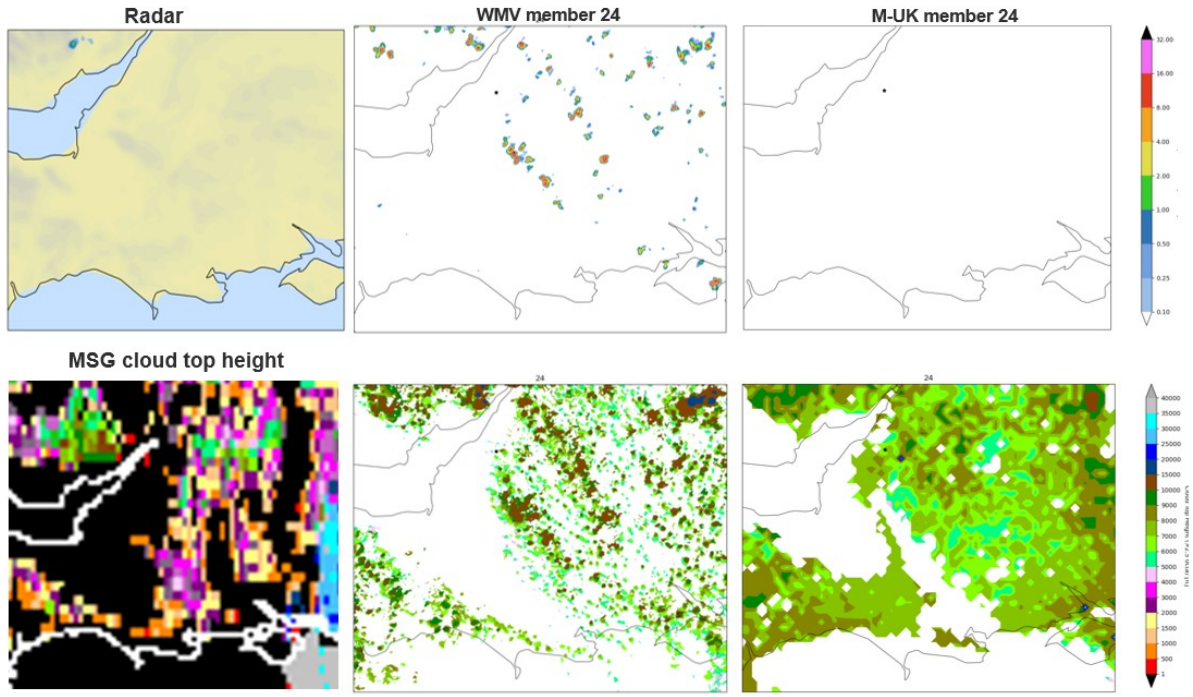


Figure 28. Surface rainfall rate from member 24 of the RAL3.1 WMV and MOGREPS-UK ensembles compared with the radar composite (top) and cloud top height from the same simulations compared with the MSG satellite cloud top height product (bottom) at 1200 UTC 22 June 2023 over the inner WMV domain.

The MSG satellite cloud top height product shows there was mostly shallow cloud (tops less than 6 kft) at this time with some deeper cloud associated with the developing shower in South Wales. Precipitating showers are generally associated with cloud tops greater than 9 kft, the WMV member which has lots of precipitating showers has deeper clouds than the corresponding MOGREPS-UK member which doesn't have any precipitating showers, although the non-precipitating clouds in MOGREPS-UK are deeper than the observed clouds. It's also noted that MOGREPS-UK has more extensive cloud coverage than the WMV, which tends to have fragmented cloud. Due to the low resolution of the

© Crown copyright 2025, Met Office

observations, it is hard to verify which model is correct. The WMV100 cloud field is similar to the WMV, i.e., it's convective cloud rather than layered cloud (Figure 29). Comparing one member with few precipitating showers (23) to a member with lots of precipitating showers (24), the member with fewer precipitating showers has fewer deep clouds than the member with more precipitating showers.

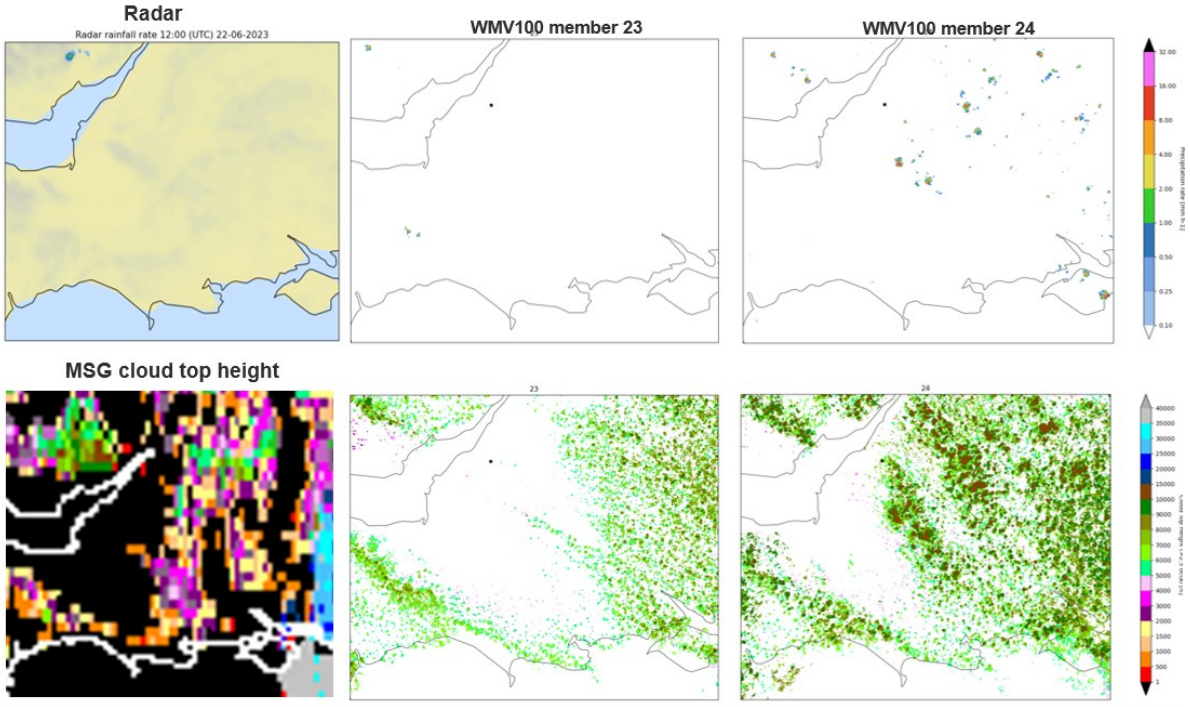


Figure 29. Surface rainfall rate from members 23 and 24 of the RAL3.1 WMV100 ensemble compared with the radar composite (top) and cloud top height from the same members compared with the MSG satellite cloud top height product (bottom) at 1200 UTC 22 June 2023 over the inner WMV domain.

This has been quantified in Figure 30, which shows the distribution of cloud top height from member 23 (few/no precipitating showers) compared to member 24 (lots of precipitating showers) for MOGREPS-UK, the WMV and WMV100.

© Crown copyright 2025, Met Office

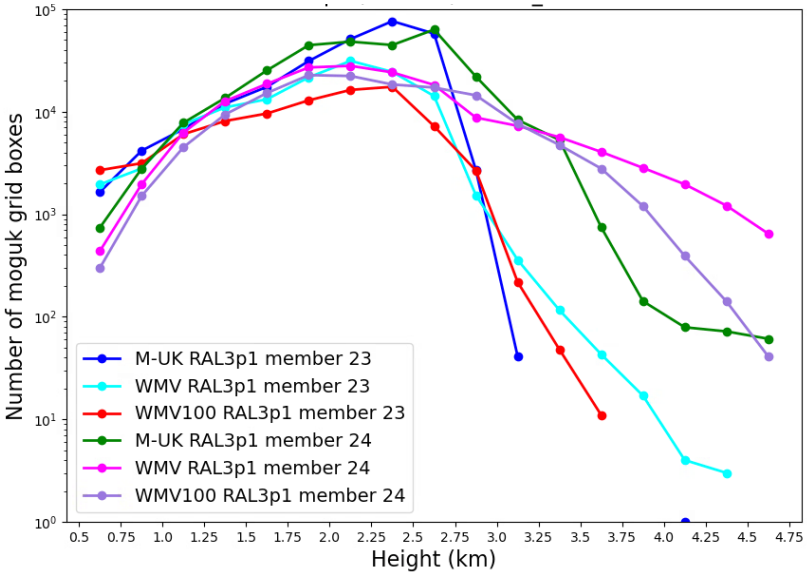


Figure 30. Distribution of cloud top height for the period 7-22 UTC 22 June 2023 from member 23 and member 24 of the RAL3.1 MOGREPS-UK, WMV and WMV100 ensembles. Data is every 5 minutes, extracted over the inner WMV domain and has been regridded to 2.2 km.

The member with more precipitating showers has more grid boxes with cloud tops > 3 km than the member with fewer precipitating showers. In both cases, the WMV has deeper clouds than the WMV100, which in turn has deeper clouds than MOGREPS-UK.

Figure 31 shows the surface-based CAPE (Convective Available Potential Energy) and CIN (Convection Inhibition) at 09 UTC on 22 June 2023 from member 24 of the MOGREPS-UK, WMV and WMV100 ensembles and member 23 of the WMV100 ensemble compared with the prototype surface meso-analysis product derived from the WOW surface station observations and MOGREPS-UK.

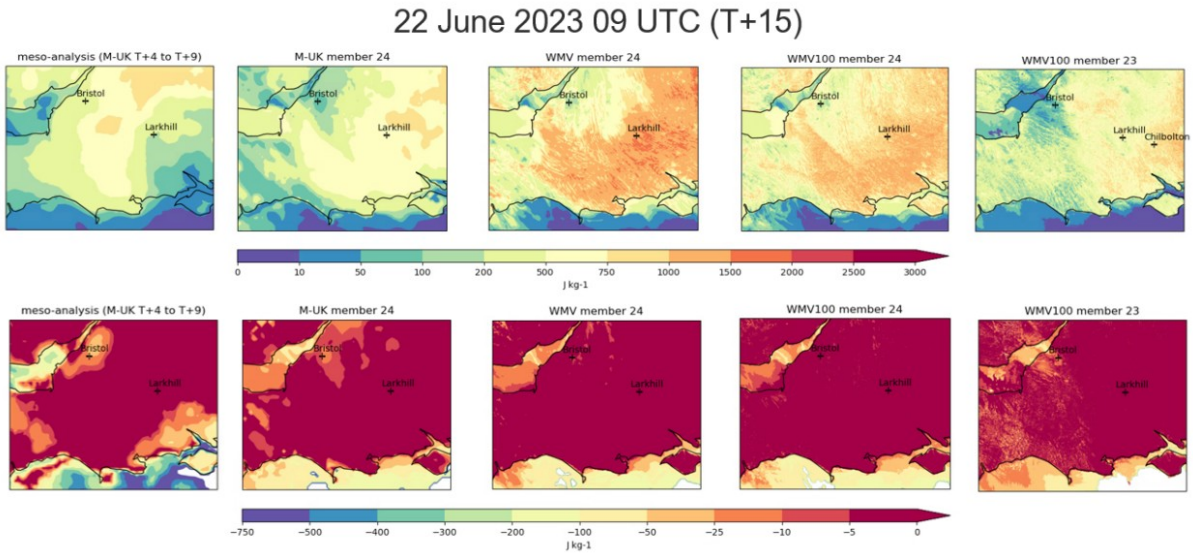


Figure 31. CAPE (top) and CIN (bottom) at 0900 UTC 22 June 2023 from member 24 of MOGREPS-UK, the WMV and WMV100 and member 23 of the WMV100 compared with the surface meso-

© Crown copyright 2025, Met Office

analysis product.

Before convection initiated, there was very little CIN in any of the models or the meso-analysis, suggesting there wasn't a particularly strong capping inversion inhibiting convection. Member 24 of the WMV has more CAPE and slightly less CIN than the driving MOGREPS-UK member (this is also true of all other members, not shown). Member 24 of the WMV100 has less CAPE than the equivalent WMV member, but more than the driving MOGREPS-UK member, which may explain why the WMV100 tends to have fewer clouds with tops exceeding 3 km and fewer precipitating showers than the WMV. Member 24 of both the WMV and WMV100 ensembles has more CAPE and less CIN than the meso-analysis. However, it should be noted that there aren't currently any direct observations of CAPE/CIN in the meso-analysis, the increments to the CAPE field are inferred from surface observations, using the covariance between the surface fields and the diagnosed surface-based CAPE/CIN values from MOGREPS-UK. Generally, the members that simulate more precipitating showers have more CAPE and slightly less CIN at 09 UTC than the members that simulate fewer precipitating showers (e.g. compare members 23 and 24 of the WMV100), suggesting increased CAPE (and slightly decreased CIN) could be one factor that allows the convective clouds to penetrate deeper leading to precipitation.

Figure 32 shows the observed and simulated atmospheric soundings at 09 UTC from Larkhill.

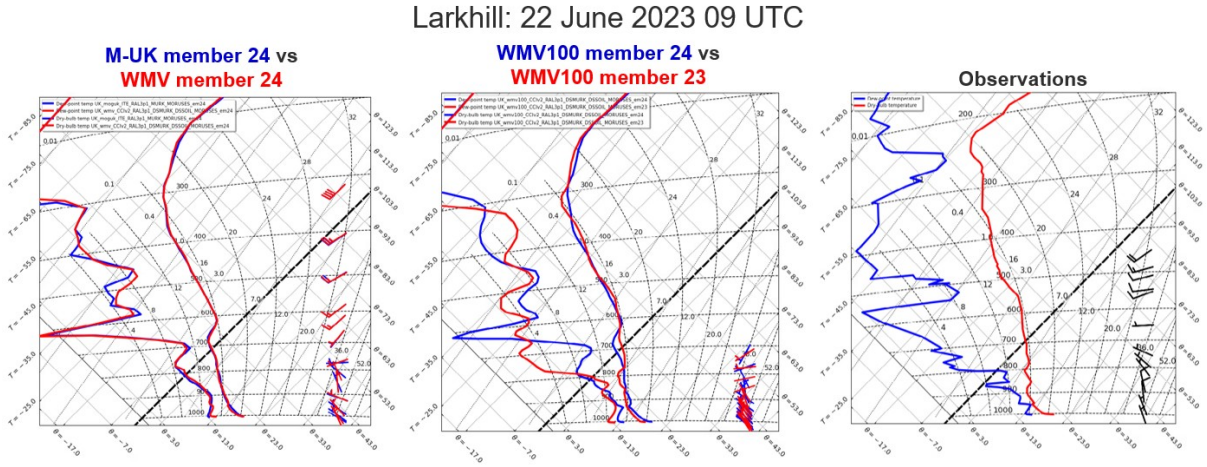


Figure 32. Atmospheric profile at Larkhill at 09 UTC 22 June 2023 from member 24 of MOGREPS-UK and the WMV (left), member 24 and member 23 of the WMV100 (middle) and the observations (right).

Unsurprisingly, the equivalent MOGREPS-UK, WMV and WMV100 members have a very similar profile, suggesting the large-scale environmental conditions are inherited from the global ensemble. MOGREPS-UK member 24 has a slightly drier (and warmer) Boundary Layer (BL) than the equivalent WMV member, which may contribute to an environment less conducive to convection. However, both

© Crown copyright 2025, Met Office

models have a drier BL than the observed sounding, even the WMV member which produces too many precipitating showers. Above the BL, member 24 is moister than the observed sounding which may be another contributing factor to why the clouds in this member are able to penetrate deeper than the observed clouds as they will be entraining less dry air. The WMV100 member with fewer showers (23) has a cooler and drier BL (possibly too dry compared to the observed sounding) than the WMV100 member with more showers (24). It is also drier aloft which better matches the observed sounding and may prevent the clouds from penetrating as deep as those in member 24 as they will be entraining more dry air.

Finally, Figure 33 shows domain-average profiles of potential temperature, specific humidity, updraft velocity (vertical velocity > 0 m/s) and cloud water (liquid + ice) at 12 UTC from member 24 of MOGREPS-UK, the WMV and WMV100 and member 23 of the WMV100.

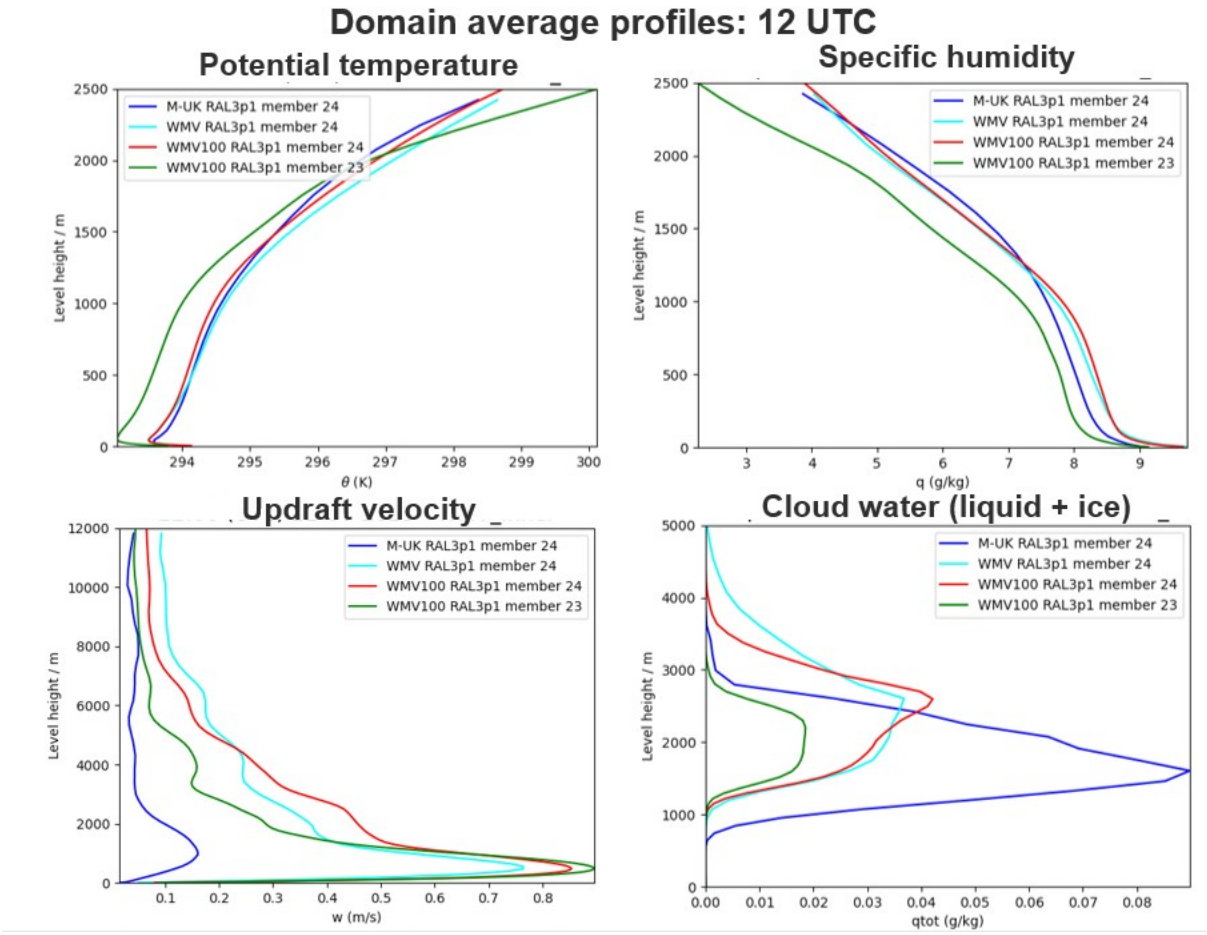


Figure 33. Domain-average profiles of potential temperature, specific humidity, updraft velocity (vertical velocity > 0 m/s) and cloud water (liquid + ice) at 1200 UTC 22 June 2023 from member 24 of MOGREPS-UK, the WMV and WMV100 and member 23 of the WMV100. Averaged over the inner WMV domain.

© Crown copyright 2025, Met Office

Member 23 (fewer showers) of the WMV100 has a cooler and drier BL than the member 24 (too many showers) simulations at all resolutions. There is little difference in potential temperature between all three member 24 simulations, however MOGREPS-UK has a drier BL than the WMV and WMV100. Consistent with Figure 30, member 24 of the WMV has deeper clouds than the WMV100, which has deeper clouds than MOGREPS-UK. MOGREPS-UK has more domain-average cloud water below 2.5 km than the WMV and WMV100, consistent with more extensive cloud coverage as noted in Figure 28. Member 23 of the WMV100 has less cloud water and shallower clouds than member 24, consistent with having fewer precipitating showers. The updraft velocity in the BL generally increases with increased resolution, however above the BL member 24 of the WMV and WMV100 both have stronger updrafts than member 23 of the WMV100. Member 23 of the WMV100 has drier air above the BL (which better matches the observed sounding) so this suggests as the updrafts entrain this dry air, they become less buoyant which prevents the clouds from growing as deep. In contrast, the updrafts in member 24 of the WMV and WMV100 are entraining less dry air above the BL so they remain more buoyant allowing the clouds to grow too deep and precipitate erroneously. Additionally, the member 24 WMV100 updrafts are stronger than the member 24 WMV updrafts up to a height of ~4 km, but above this they are weaker suggesting the updrafts in the WMV100 simulations don’t go as high, which may explain why the WMV100 clouds are shallower than the clouds in the equivalent WMV member. A possible reason for this is that the WMV100 updrafts are entraining more environmental air as the turbulence is better resolved.

6. Conclusions

During summer 2023 evaluation of the WMV was conducted through a daily running trial, running as an ensemble nested inside the operational MOGREPS-UK 2.2km ensemble. Overall, the WMV looks promising for high-impact convective events as it is better able to represent the storm structure and organisation, which is often lacking in MOGREPS-UK. This often leads to more reliable probabilities of heavy rainfall in the WMV ensemble compared to MOGREPS-UK. The 2023 UK Summer Testbed also highlighted that the WMV probabilities of heavy precipitation were more often favoured over MOGREPS-UK (Flynn et al., 2023). In several cases very high 10m wind gusts were associated with the convective precipitation in the WMV, further work and comparison with bespoke observations (such as those collected during WesCon and the ASSURE campaign in Bristol) is required to assess how realistic these are.

© Crown copyright 2025, Met Office

Fraction Skill Score analysis showed that the WMV is more skilful at smaller neighbourhoods than MOGREPS-UK with the displacement between observed and modelled precipitation being less in the WMV than in MOGREPS-UK. However, the WMV ensemble is under-spread compared to MOGREPS-UK.

While the WMV performs well in organised and deep convective cases, there is still an issue with the WMV, and urban-scale models in general, producing too many small precipitating showers in situations where there should only be shallow clouds. This is thought to be a result of shallow clouds getting too deep in the model and precipitating erroneously. Decreasing the horizontal grid length somewhat improves this issue, however it does not solve it entirely. At 100m grid length the updrafts (and hence the clouds) do not penetrate as deep, despite being stronger in the BL, leading to less precipitation. It is hypothesised that the entrainment processes are better resolved at 100m resulting in more entrainment of dry air and less buoyant updrafts. Further work is required to fully understand the cause of this issue with the WesCon observations expected to aid this strand of work. Preliminary comparisons of the model vertical velocities with the FAAM aircraft observations show the distribution of updraft speeds in the 100m model compare best to the observed distribution, with MOGREPS-UK lying furthest from the observations. Comparing the model thermodynamic profiles against sondes around the time of the showers could help determine whether errors in the urban-scale models are due to the large-scale state being wrong in the initial conditions and LBCs; if the profiles are just slightly too unstable or moist compared to observations, that might explain the clouds going too deep. Comparing the profiles from one ensemble member that had a lot of excessive showers with a member that had fewer showers (closer matching the radar observations) showed that compared to observations the member with more showers was too moist above the BL, with this error inherited from MOGREPS-UK (likely inherited from the Global ensemble). In reality, entrainment of dry air above the BL may have inhibited deep convection resulting in shallow clouds, however the WMV and WMV100 members that missed the dry layer were entraining moister air resulting in more buoyant updrafts and deeper clouds. This highlights the benefit of using ensembles for process studies as comparing “good” and “bad” ensemble members can provide physical insight into the mechanisms leading to convection.

The ability to run equivalent UM and LFRic simulations will isolate whether the excessive small showers problem is due to the physics or the dynamical core. Microphysical sensitivity could also be a contributing factor. We know the clouds are going too deep but as there’s a too-way interaction between microphysics and dynamics, it’s not straightforward to separate cause and effect (e.g., the clouds could be raining-out cloud-water too readily, which results in a loss of water-loading, making the clouds more buoyant so they go deeper).

© Crown copyright 2025, Met Office

Acknowledgments

Thanks to Anne McCabe for help and assistance with running the Ensemble Nesting Suite in real-time and for providing the ensemble visualisation suite and code for calculating the dFSS/eFFS metrics. Additionally, thanks to Stu Webster for making the WMV output available to external WesCon and UK Testbed participants.

© Crown copyright 2025, Met Office

References

Boutle IA, Finnenkoetter A, Lock AP, Wells H. (2016). The London Model: forecasting fog at 333 m resolution. *Q. J. R. Meteorol. Soc., 142*, https://doi.org/10.1002/qj.2656.

Bush M, Allen T, Bain C, Boutle I, Edwards J, Finnenkoetter A, Franklin C, Hanley K, Lean H, Lock A, Manners J, Mittermaier M, Morcrette C, North R, Petch J, Short C, Vosper S, Walters D, Webster S, Weeks M, Wilkinson J, Wood N, Zerroukat M. (2020). The first Met Office Unified Model-JULES Regional Atmosphere and Land configuration, RAL1. *Geosci. Model Dev., 13*, 1999-2029, https://doi.org/10.5194/gmd-13-1999-2020.

Bush M, Boutle I, Edwards J, Finnenkoetter A, Franklin C, Hanley K, Jayakumar A, Lewis H, Lock A, Mittermaier M, Mohandas S, North R, Porson A, Roux B, Webster S, Weeks M. (2023). The second Met Office Unified Model/JULES Regional Atmosphere and Land configuration, RAL2. *Geosci. Model Dev.*, https://doi.org/10.5194/gmd-2022-209.

Bush M, Flack D, Coauthors. (2024). Unifying Mid-latitude and Tropical Regional Model Configurations: The third Met Office Unified Model–JULES Regional Atmosphere and Land Configuration, RAL3. *Geosci. Model Dev.*, https://doi.org/10.5194/gmd-2024-201.

Dey SRA, Leoncini G, Roberts NM, Plant RS, Migliorini S. (2014). A spatial view of ensemble spread in convection-permitting ensembles. *Monthly Weather Review, 142*, 4091–4107, https://doi.org/10.1175/MWR-D-14-00172.1.

Dey SRA, Plant RS, Roberts NM, Migliorini S. (2016a). Assessing spatial precipitation uncertainties in a convective-scale ensemble. *Q. J. R. Meteorol. Soc., 142*, 2935-2948, https://doi.org/10.1002/qj.2893.

Field PR, Hill A, Shipway B, Furtado K, Wilkinson J, Miltenberger A, Gordon H, Grosvenor DP, Stevens R, Van Weverberg K. (2023). Implementation of a Double Moment Cloud Microphysics Scheme in the UK Met Office Regional Numerical Weather Prediction Model. *Q. J. R. Meteorol. Soc.*, https://doi.org/10.1002/qj.4414.

Flynn A, Hanley K, Porson A, Lean H. (2023). Post-event Sub-km Ensemble UK Testbed 2023: Focus on the convection characteristics. *Met Office Technical Report*.

Furtado K, Field PR, Boutle IA, Morcrette CJ, Wilkinson JM. (2016). A physically based subgrid parameterization for the production and maintenance of mixed-phase clouds in a general circulation model. *J. Atmos. Sci., 73*, 279–

© Crown copyright 2025, Met Office

291.

Glazer R, Flynn A, Porson A, Grimmond S. (2023). UK Testbed 2023: Post-event Sub-km Ensemble and Urban-scale Evaluation. *Met Office Technical Report*.

Hanley KE, Lean HW. (2024). The performance of a variable resolution 300m ensemble for forecasting convection over London. *Q. J. R. Meteorol. Soc., 150*, https://doi.org/10.1002/qj.4794.

Hanley KE, Plant RS, Stein THM, Hogan RJ, Nicol JC, Lean HW, Halliwell C, Clark PA. (2015). Mixing length controls on high resolution simulations of convective storms. *Q. J. R. Meteorol. Soc., 141*, DOI:10.1002/qj.2356.

Harrison DL, Norman K, Pierce C, Gaussiat N. (2011). Radar products for hydrological applications in the UK. *Proceedings of the Institution of Civil Engineers, 165*, 89-103.

Lorenz EN. (1969). The predictability of a flow which possesses many scales of motion. *Tellus, 21A*, 289–307, doi:10.3402/tellusa.v21i3.10086.

McCaul EW, Goodman SJ, LaCasse KM, Cecil DJ. (2009). Lightning threat using cloud-resolving model simulations. *Wea. Forecasting, 24*, 709-729, doi:10.1175/2008WAF2222152.1.

Porson A, Clark PA, Harman IN, Best MJ, Belcher SE. (2010). Implementation of a new urban energy budget scheme in the MetUM. Part I: Description and idealized simulations. *Q. J. R. Meteorol. Soc., 136*, 1514-1529, https://doi.org/10.1002/qj.668.

Porson AN, Carr JM, Hagelin S, Darvell R, North R, Walters D, Mylne KR, Mittermaier MP, Willington S, Macpherson B. (2020). Recent upgrades to the Met Office convective-scale ensemble: An hourly time-lagged 5-day ensemble. *Q. J. R. Meteorol. Soc., 146*, https://doi.org/10.1002/qj.3844.

Roberts NM, Lean HW. (2008). Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review, 136*, 78-97, https://doi.org/10.1175/2007MWR2123.1.

Schwartz CS and Sobash RA. (2017). Generating Probabilistic Forecasts from Convection-Allowing Ensembles Using Neighborhood Approaches: A Review and Recommendations. *Mon. Wea. Rev., 145*, 3397–3418, https://doi.org/10.1175/MWR-D-16-0400.1. .

Skok G. (2015). Analysis of Fraction Skill Score properties for a displaced rainband in a rectangular domain. *Meteorol. Appl., 22*, 477–484.

Skok G. (2016). Analysis of Fraction Skill Score properties for a displaced rainy grid point in a rectangular domain. *Atmos. Res., 169*, 556–565.

© Crown copyright 2025, Met Office

Skok G, Roberts N. (2017). Estimating the displacement in precipitation forecasts using the Fractions Skill Score. *Quarterly Journal of the Royal Meteorological Society, 144*, 414-425. doi:https://doi.org/10.1002/qj.3212

Smith RNB. (1990). A scheme for predicting layer clouds and their water contents in a general circulation model. *Q. J. R. Meteorol. Soc, 116*, 435–460.

Van Weverberg K, Morcrette CJ, Boutle I, Furtado K, Field PR. (2021). A Bimodal Diagnostic Cloud Fraction Parameterization. Part I: Motivating Analysis and Scheme Description. *Mon. Wea. Rev., 149*, 841–857.

Wilson DR, Ballard SP. (1999). A microphysically based precipitation scheme for the UK Meteorological Office Unified Model. *Q. J. R. Meteorol. Soc., 125*, 1607-1636.

© Crown copyright 2025, Met Office